dc_52_10

Geometriai Problémák az Addititív Kombinatorikában

Solymosi József

Akadémiai doktori értekezés

1 Bevezető

A doktori fokozat megszerzése óta az additív kombinatorikával és ahhoz kapcsolodó problémákkal foglalkozom. Ennek a fiatal matematikai kutatási területnek számos ága van fontos alkalmazásokkal a számítógéptudományban, számelméletben, kombinatorikában vagy éppen harmonikus analízisben. Különösen kedvelem azokat a problémákat ahol közvetve vagy közvetlenül geometriát lehet alkalmazni számelméleti kérdésekben, illetve amikor számelméleteti eredmények segítségevel oldhatóak meg geometriai problémák. Kutatásom három – egymással kapcsolatban álló – témakör köré csoportosítható: Szemerédi tételének általánosításai, Erdős és Szemerédi összeg-szorzat sejtése és Erdős különböző távolságokkal kapcsolatos sejtései.

Jelen értekezésben mindhárom témakörben bemutatunk eredményeket. Tizenkét cikket fogunk ismertetni négy fejezetben a következők szerint:

1. Első fejezet: A "Hypergraph Removal Lemma" alkalmazásai a Szemerédi tétel általánosításában.
   a. Bevezetés: Regularity, uniformity, and quasirandomness
   b. Első cikk: Roth tételének általánosítása.
   c. Második cikk: Erdős és Graham egy probléméjáról.
   d. Harmadik cikk: Számtani sorozatok kis összegű halmazokban
2. Második fejezet: Az összeg-szorzat problémáról
   e. Negyedik cikk: Összeg-szorzat becslés komplex számokra
   f. Ötödik cikk: Javított becslés a Szemerédi-Trotter tétel alkalmazásával
   g. Hatodik cikk: Összeg-szorzat becslés mátrixokra
   h. Hetedik cikk: További javítás elemi lineáris algebra alkalmazásával
3. Harmadik fejezet: Számelméleti eredmények alkalmazása a diszkrét geometria területén.
   i. Nyolcadik cikk: Extremális pont-egyenes illeszkedési rendszerek lokális struktúrája
   j. Kilencedik cikk: Összeg-szorzat becslések geometriai alkalmazása
   k. Tizedik cikk: Erdős és Ulam egy problémájáról
4. Negyedik fejezet: Különböző távolságok
   l. Tizenegyedik cikk: A különböző távolságok száma magasabb dimenzióban
   m. Tizenkettedik cikk: Különböző távolságok homogén ponthalmazokban

2 Szemerédi tételének általánosítása

Szemerédi Endre bizonyította 1970-ben Erdős és Turán sejtését miszerint az egészek bármely sűrű részhalmaza tartalmaz tetszőlegesen hosszú számtani sorozatokat. Ezt a fontos eredményt ergodikus módszerekkel újrabizonyította és általánosította Fürstenberg és Katznelson [7].

Megmutatták hogy bármely $\delta > 0$ és pozitív egész r-re minden $X \subset \mathbf{Z}^r$ halmazhoz van olyan N hogy minden $A \subset \{1, 2, \ldots, N\}^r$ esetén ha $|A| \geq \delta N^r$ akkor A-ban talalható egy részhalmaz ami $a + dX$ alaku. (d egy pozitív egész)

2000-ben Timothy Gowers egy analitikus bizonyítást dolgozott ki Szemerédi tételére. Roth 3-hosszú számtani sorozatokra vonatkozó tételénél alkalmazott technikát terjesztette ki a hosszabb számtani sorozatok problémájára.

A publikáció előtti kézirat végén Gowers fontos nyitott problémának jelölte meg hogy találjunk egy elemi bizonyítást Roth tételének egy kétdimenziós változatára (a kérdés részleteit hamarosan meglátjuk). A kézirat olvasása után egy egyszerű bizonyítást találtam Szemerédi és Ruzsa "6;3" tétele alkalmazásával. Az itt alkalmazott módszer – amit később általánosítottam – lehetőséget adott Fürstenberg és Katznelson fent említett tételének elemi bizonyítására. (egy másik, sokkal nehezebb, gráfelméleti eredmény segítségével)

[15]-ben bizonyítottam hogy minden sűrű részhalmaza a kétdimenziós egész rácsnak tartalmaz egy négyzetet. (A Fürstenberg-Katznelson tétel speciális este amikor $X = \{(0,0),(0,1),(1,0),(1,1)\}$) Azt is megmutattam, hogy a Fürstenberg-Katznelson tétel következik egy – akkor még csak sejtett – állításból, az úgynevezett "Hypergraph Removal Lemma"-ból.

A Hypergraph Removal Lemma következik Szemerédi gráf-regularitási lemmájának általánosításából [1] ; a hipergráf regularitási lemma és az ehhez tartozó leszámolási lemma alkalmazásából. Ezeket az állításokat egymástól függetlenül Tim Gowers és Vojta Rödl diákjaival igazolták. (Gowers [3] and Rödl et al. [2])

Rödl, Nagle, Skokan, Schacht és Kohayakava cikke, "The hypergraph regularity method and its applications" a Proceedings of the National Academy of Sciences of USA-ben jelent meg. Az újság szerkesztői felkértek, hogy írjak egy "Commentary"-t a cikkhez, amit csak az általános tudomány kiemelt fontosságú eredményekhez szoktak kérni. Ezt az írást is csatoltam a doktori dolgozatomhoz, bevezetőként az első fejezethez.

A Hipergráf Regularitási Lemma az első két bizonyítás után újabb bizonyításokat

[1] A Regularitási Lemma a diszkrét matematika egyik legfontosabb eszköze. Szemerédi a fentiekben már említett Erdős-Turán sejtés bizonyításához fejlesztette ki, mely szerint egy pozitív felső sűrűségű egész számokból alló sorozat tartalmaz hosszú számtani sorozatokat [32]. A lemmát úgy lehetne röviden összefoglalni, hogy bizonyos értelemben minden nagy gráfot jól lehet közelíteni kisebb súlyozott élű gráfokkal. A Regularitási Lemmával és az ezzel a módszerrel kapcsolatos további információkért lásd a [31] áttekintő cikket és a további regularitásra vonatkozó referenciákat a cikkjegyzékben

kapott, Terry Tao, Elek Gábor és Szegedi Balázs, valamint Yoshi Ishigami is bizonyította, részben a korábbi bizonyításokra alapozva. (Vannak akik kétlkednek Ishigami bizonyításának korrektségében) Bár ez az eredmény túl mutat jelen doktori dolgozat keretein, megemlítjük még, hogy Szemerédi tételét is használva és részben a hipergráf regularitási lemma által inspirálva Ben Green és Terry Tao bebizonyította hogy a prímszámok között tetszőlegesen hosszú számtani sorozatok talalhatók.

3 Az összeg-szorzat probléma

Minden olyan probléma ide sorolható ami a két művelet, az összeadás és a szorzás összeférhetetlenségét mutatja; Ha egy halmaz összeghalmaza nem sokkal nagyobb mint az eredeti halmaz akkor a szorzathalmaz nagy kell hogy legyen. Ezt az állítást pontosan megfogalmazzuk valós számok véges részhalmazaira.

Legyen A valós számok egy véges részhalmaza. Az összegehalazt az alábbiak szerint definiáljuk:

$$A + A = \{a + b | a, b \in A\}.$$

Hasonlóan, a szorzathalmazt a következőképpen kapjuk:

$$A \cdot A = \{ab | a, b \in A\}.$$

Erdős és Szemerédi azt sejtette hogy az összeghalmaz vagy a szorzathalmaz mindig majdnem kvadratikus méretben az eredi halmazhoz képest.

$$\max(|A + A|, |A \cdot A|) \geq |A|^{2-\delta}$$

ahol δ tart nullához amint $|A|$ tart a végtelenhez.

Egy rendkívül elegáns cikkben Elekes [26] megmutatta, hogy diszkrét geometria használható jó összeg-szorzat becslésekhez. Elkes módszerét továbbfejlesztve megmutattam [10]-ban, hogy

$$\max(|A + A|, |A \cdot A|) \geq c|A|^{14/11}/\log|A|.$$

Egy Tardos Gáborral közösen írt cikkben igazoltuk, hogy a fenti egyenlőtlenség komplex számok véges halmazára is igaz. Ezzel megjavítottuk egy korábbi eredményemet ahol a

$$\max(|A + A|, |A \cdot A|) \geq c|A|^{5/4}$$

egyenlőtlenséget igazoltam komplex számokra. Mindamellett, az ott alkalmazott módszerem általánosítható más testek/gyűrűk feletti összeg-szorzat problémákra is, így ez az eredmény bekerült több egyetemi jegyzetbe.

A közelmúltban egy még egyszerűbb bizonyítást találtam az erősebb

$$\max(|A + A|, |A \cdot A|) \geq c|A|^{4/3}\log|A|$$

egyenlőtlenségre amikor A valós számok részhalmaza [11].

Van Vu-val közös cikkben a négyzetes matrixok gyűrűje felett igazoltuk a

$$\max(|A+A|, |A \cdot A|) \geq c|A|^{5/4}$$

egyenlőtlenséget "szép" mátrixok családjára.

Az összeg-szorzat probléma nagyon érdekes és fontos alkalmazásokkal bír a véges testek felett is. Itt persze további megkötésekre van szükség hiszen például egy részgyűrűnek az összeghalmaza és a szorzathalmaza is nagyon kicsi, a fenti egyenlőtlenségekhez hasnonlóak általánosan nem várhatóak.

Bourgain, Katz és Tao bizonyított egy $|A|^{1+\varepsilon}$ alsó becslést a véges testek felett [24] az alábbiak szerint: Legyen $A \subset \mathbf{F}_p$ és $p^\alpha \leq |A| \leq p^{1-\alpha}$. Ekkor van olyan $\varepsilon > 0$ ami csak α-tól függ hogy

$$\max(|A+A|, |A \cdot A|) \geq c|A|^{1+\varepsilon}.$$

Ez az eredmény fontos alkalmazásokkal bír számelméletben, számítógéptudományban, Ramsey elméletben és kriptográfiában. Kiderült, hogy a valós esetre használt geometriai látásmód a véges karakterisztikájú testek felett is alkalmazható. Hart és Iosevich-el közös cikkünkben [25] elsőként adtunk jó becslést $\max(|A+A|, |A \cdot A|)$-re ahol $A \subset \mathbf{F}_p$ and $p^{1/2} \ll |A| \ll p$. (Hozzá kell tennem, hogy a legtöbb alkalmazáshoz a $|A| \ll p^{1/2}$ szakasz az érdekes)

4 Különböző távolságok

Ez a harmadik témakör érdekesen kapcsolódik az additív kombinatorikához. A korábban emített Bourgain, Katz, Tao cikk foglalkozik a különböző távolságok problémájával véges testek felett. Megmutatták hogy a probléma bizonyos értelemben ekvivalens az összeg-szorzat kéréssel $\mathbf{F}_p^2$-ben.

Először Erdős egy klasszikus problémáját tárgyaljuk. Erdős írja [28]-ben: "My most striking contribution to geometry is, no doubt, my problem on the number of distinct distances."

Jelölje $g(n)$ egy n-elemű síkbeli ponthalmaz által meghatározott különböző távolságok lehetséges minimumális számát. Erdős megmutatta, hogy a $\sqrt{n} \times \sqrt{n}$ méretű egész rács pontjai $cn/\sqrt{\log n}$ különböző távolságot határoznak meg. Úgy sejtette, hogy hasonló becslés felülről is igaz.

Tóth Csabával [13]-ben megmutattuk hogy $g(n) > cn^{6/7}$. Székely immár klasszikusnak mondható módszerét javítottuk meg. Cikkünk egy számelméleti lemmáját Katz és Tardos megjavították, megnövelve a 6/7 kitevőt egy további 0.007-tel. Ez a mostani rekord, és Ruzsa Imre egy konstrukcióval megmutatta hogy jelen módszerünkkel Erdős sejtett becslése nem elérhető, további, új ötletekre lesz szükség az előrelépéshez.

Jelen dolgozatban a magasabb dimenziós változatát vizsgáljuk Erdős sejtésének. A sejtés (Erdős) szerint n pont a d-dimenziós euklideszi térben legalább $n^{\frac{2}{d}-\epsilon}$ különböző távolságot határoz meg.

Van Vu-val közös cikkünkben [16] elsőnek sikerült megmutatnunk hogy Erdős sejtése asszimptotikusan igaz;

n pont a d-dimenziós euklideszi térben legalább

$$n^{\frac{2}{d}-\frac{2}{d(d+2)}}$$

különböző távolságot határoz meg.

Harmonikus analizisben kutatókat érdekli a különböző távolságok probléma egyenletes eloszlású ponthalmazokra is, ahol esetleg jobb becslés várható. Tom Wolff munkássága alapján Łaba, Iosevich és mások is kapcsolatot találtak a híres Kakeya sejtés és a különböző távolságok problémája egyenletes eloszlású ponthalmazokra között.

Tóth Csabával közös cikkünkben [19] az erősebb

$$n^{\frac{2d}{d^2+1}}$$

alsó becslést bizonyítottuk egyenletes eloszlású ponthalmazokra.

5 A cikkek bemutatása

Ebben az összefoglaló szekcióban röviden bemutatjuk a tézis cikkeit. A bemutatás során használjuk a jelöléseket az előző bekezdésekből.

1. Első fejezet: A "Hypergraph Removal Lemma" alkalmazásai a Szemerédi tétel általánosításában.

   a. Bevezetés: Regularity, uniformity, and quasirandomness

   [20]'Regularity, uniformity, and quasirandomness'. Proceedings of the National Academy of Sciences of the United States of America. 102.23 (2005): 8075 - 8076.

   Ezt a cikket bevezetőnek szántam az első fejezethez. Új önálló eredményt nem tartalmaz, de segít a későbbi erdmények megértésében.

   b. **Első cikk:** Roth tételének általánosítása.

   [21]'Note on a generalization of Roth's theorem'. Discrete and computational geometry; Algorithms Combin. Vol. 25. Ed. Janos Pach. Springer, 2003. 825 – 827.

   Gowers kérdésére válaszolva egyszerű bizonyítást adunk a következő problémára:

Bármely $\delta > 0$-hoz van olyan N hogy minden $A \subset \{1, 2, \ldots, N\}$ esetén ha $|A| \geq \delta N^2$ akkor A-ban talalható három pont amik egy derékszögű egyenlőszárú háromszöget alkotnak, azaz $(x, y), (x+d, y), (x, y+d)$ alakuak. (d egy nemnulla egész)

Ezt az eredményt nemrég Ilya Shkredov megjavította. Továbbfejlesztve Gowers és Bourgain analitikus módszereit megmutatta hogy $(\log\log\log n)^{-1}$ sűrűség garantál ilyen háromszöget. (Az én bizonyításom csak $(\log^* n)^{-1}$ sűrűségre működik)

c. **Második cikk:** Erdős és Graham egy problémájáról.

[15] 'A note on a question of Erdős and Graham', Combin. Probab. Comput. 13 (2004), no. 2, 263–267.

A fő erdménye a cikknek egy új módszer bevezetése; hogyan használható a "Removal Lemma" a többdimenziós Szemerédi tétel bizonyítására.

Példaként elemi bizonyítást adtunk Erdős és Graham egy kérdésére, miszerint minden sűrű részhalmaza a kétdimenziós egész rácsnak tartalmaz egy négyzetet. Ez a Fürstenberg-Katznelson tétel speciális este amikor

$X = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$

d. **Harmadik cikk:** Számtani sorozatok kis összegű halmazokban

[18] 'Arithmetic Progressions in Sets with Small Sumsets', Combinatorics, Probability and Computing. 15 (2006): 597 - 603.

Ez a cikk egy további illusztráció a "Removal Lemma" erjére. Megmutattuk, hogy ha $|A + A| \leq C|A|$ akkor A tartalmaz hosszú számtani sorozatokat. Ezt akkor is meg tudjuk mutatni, ha az összeghalmaz csak egy sűrű gráf mentén kicsi. Balog és Szemerédi [1] egy tétele alapján tudjuk, hogy ez az eset visszavezethető az előzőre, de itt nem kell használnunk ezt az eredményt. A fő érdekesség azonban nem ez, hanem hogy bizonyítani tudjuk a fenti állítást a nehéz Freiman-Ruzsa tétel [4] alkalmazása nélkül is.

2. Második fejezet: Az összeg-szorzat problémáról

e. **Negyedik cikk:** Összeg-szorzat becslés komplex számokra

A

$$\max(|A + A|, |A \cdot A|) \geq c|A|^{5/4}$$

egyenlőtlenséget igazoljuk komplex számokra. Az itt alkalmazott módszer általánosítható más testek/gyűrűk feletti összeg-szorzat problémákra is.

f. **Ötödik cikk:** Javított becslés a Szemerédi-Trotter tétel alkalmazásával

[10] 'On the number of sums and products', Bull. London Math. Soc. 37 (2005), no. 4, 491–494.

Elekes ötletét továbbfejlesztve igazoljuk az alábbi összeg-szorzat becslést:

$$\max(|A+A|, |A\cdot A|) \geq c|A|^{14/11}/\log|A|.$$

Mint Elekesnél is, a bizonyítás fő eleme Szemerédi és Trotter becslése egyenesek és pontok illeszkedésére.

g. **Hatodik cikk:** Összeg-szorzat becslés mátrixokra

[17](Van Vu-val közös cikk) 'Sum-product estimates for well-conditioned matrices'. Bulletin of the London Mathematical Society 2009 41(5):817-822

a négyzetes matrixok gyűrűje felett igazoltuk a

$$\max(|A+A|, |A\cdot A|) \geq c|A|^{5/4}$$

egyenlőtlenséget "well-conditioned" mátrixok családjára, azaz olyan mátrixokra amelyeknek a legnagyobb és legkisebb sajátértékei hányadosa nem túl nagy.

h. **Hetedik cikk:** További javítás elemi lineáris algebra alkalmazásával

[11] 'Bounding multiplicative energy by the sumset', Advances in Mathematics, Volume 222, Issue 2, 2009, 402–408.

Igazoljuk a

$$\max(|A+A|, |A\cdot A|) \geq c|A|^{4/3}\log|A|$$

egyenlőtlenséget a valós számok egy A részhalmazára.

3. Harmadik fejezet: Számelméleti eredmények alkalmazása a diszkrét geometria területén.

i. **Nyolcadik cikk:** Extremális pont-egyenes illeszkedési rendszerek lokális struktúrája

[12]'Dense arrangements are locally very dense I.'. SIAM JOURNAL ON DISCRETE MATHEMATICS. 20.3 (2006): 623 - 627.

Olyan pont-egyenes rendszerek szerkezetét vizsgáljuk ahol az illeszkedések száma közel van a Szemerédi-Trotter becslés által adott korláthoz. Megmutatjuk – ami intuitive sejthető – hogy az ilyen rendszerek tartalmaznak háromszögeket,

sőt nagyobb teljes részstruktúrákat is. Ez az első ilyen struktúra erdmény. A bizonyítás Szemerédi regularitási lemmáján alapul, illetve Ruzsa-Szemerédi tételtét használja.

j. **Kilencedik cikk:** Összeg-szorzat becslések geometriai alkalmazása

(Mei-Chu Changgal közös cikk) 'Sum-product theorems and incidence geometry'. JOURNAL OF THE EUROPEAN MATHEMATICAL SOCIETY. 9.3 (2007): 545 - 560.

Összeg-szorzat becsléseket alkalmazunk geometriai állítások igazolására. Egy tipikus állítás a következő: Ha a síkban négy ponton keresztül úgy adott $n-n-n-n$ egyenes, hogy legalább $n^{1.9}$ pont illeszkedik négy egyenesre, akkor a négy pont kollineáris. A bizonyításra az adott lehetőséget, hogy az összeg-szorzat becsléseknél alkalmazott geometriai technikák általánosan is alkalmazhatók.

k. **Tizedik cikk:** Erdős és Ulam egy problémájáról

[23] (Frank De Zeeuw-al közös cikk) 'On a question of Erdos and Ulam'. Discrete and Computational Geometry, Volume 43, Issue 2 (2010), Page 393-401.

Erdős kérdezte hogy vajon bármely k természetes számra megadható-e k általános helyzetű pont a síkon (nincs három egy egyenesen és négy egy körön) úgy hogy bármely kettő távolsága egész szám? Sasha Kurz talált egy ilyen egész távolságú ponthalmazt hét ponton, ez eddig a rekord. A másik oldalról Ulam kérdezte, hogy megadható-e egy mindenütt sűrű ponthalmaz, hogy bármely kettő távolsága racionális szám. Erdős sejtette hogy ez nem lehetséges. Több kutató is próbálkozott racionális távolságú ponthalmazokat találni algebrai görbék mentén.

Diákommal, Frank De Zeeuw-val, megmutattuk hogy ha egy algebrai görbe tartalmaz egy végtelen racionális ponthalmazt, akkor a pontok egy körön vagy egyenesen vannak.

4. Negyedik fejezet: Különböző távolságok

l. **Tizenegyedik cikk:** A különböző távolságok száma magasabb dimenzióban

[16] (Van Vu-val közös cikk) Near optimal bound for the distinct distances problem in high dimensions. COMBINATORICA, 28.1 (2008): 113 – 125.

Erdős sejtésének megfelelően n pont a d-dimenziós euklideszi térben legalább

$$n^{\frac{2}{d}-\frac{2}{d(d+2)}}$$

különböző távolságot határoz meg.

m. **Tizenkettedik cikk:** Különböző távolságok homogén ponthalmazokban

[19](Tóth, Csabával közös cikk) Distinct distances in homogeneous sets in Euclidean space. Discrete Comput. Geom. 35 (2006), no. 4, 537–549.

Ha n pont egyenletes eloszlású a d-dimenziós euklideszi térben akkor ezek legalább

$$n^{\frac{2d}{d^2+1}}$$

különböző távolságot határoznak meg.

References

[1] A. Balog, E. Szemerédi, A statistical theorem of set addition, Combinatorica 14 (1994), 263268.

[2] Rödl, V., Nagle, B., Skokan, J., Schacht, M., Kohayakava, Y. The hypergraph regularity method and its applications. Proc. Natl. Acad. Sci. USA 102 (2005), no. 23, 8109–8113

[3] Gowers, W. T. Quasirandomness, counting and regularity for 3-uniform hypergraphs. Combin. Probab. Comput. 15 (2006), no. 1-2, 143–184.

[4] I. Ruzsa, Generalized arithmetic progressions and sumsets, Acta Math. Hungar. 65 (1994), 379388.

[5] Szemerédi, E. On sets of integers containing no k elements in arithmetic progression. Collection of articles in memory of JuriĭVladimirovič Linnik. Acta Arith. 27 (1975), 199–245.

[6] Gowers, W. T. A new proof of Szemerédi's theorem, (2001) *Geom. Funct. Anal.* **11**, 465–588.

[7] Furstenberg, H. Katznelson, Y. (1978) *J. Analyse Math.* **34**, 275–291.

[8] Furstenberg, H.; Katznelson, Y. A density version of the Hales-Jewett theorem. J. Anal. Math. 57 (1991), 64–119.

[9] Furstenberg, H.; Katznelson, Y. A density version of the Hales-Jewett theorem for $k = 3$. Graph theory and combinatorics (Cambridge, 1988). Discrete Math. 75 (1989), no. 1-3, 227–241.

[10] Solymosi, Jozsef, On the number of sums and products. Bull. London Math. Soc. 37 (2005), no. 4, 491–494.

[11] Solymosi, Jozsef, Bounding multiplicative energy by the sumset, Advances in Mathematics, Volume 222, Issue 2, 2009, 402–408

[12] Solymosi, Jozsef. 'Dense arrangements are locally very dense I.'. SIAM JOURNAL ON DISCRETE MATHEMATICS. 20.3 (2006): 623 - 627.

[13] Solymosi, J.; Tóth, Cs. D. Distinct distances in the plane. Discrete Comput. Geom. 25 (2001), no. 4, 629–634.

[14] Solymosi, Jozsef; Tardos, Gábor; Tóth, Csaba D. The k most frequent distances in the plane. Discrete Comput. Geom. 28 (2002), no. 4, 639–648.

[15] Solymosi, J. A note on a question of Erdős and Graham. Combin. Probab. Comput. 13 (2004), no. 2, 263–267.

[16] Solymosi, Jozsef; Van Vu. Near optimal bound for the distinct distances problem in high dimensions. COMBINATORICA, 28.1 (2008): 113 – 125.

[17] Solymosi, Jozsef and Van Vu. 'Sum-product estimates for well-conditioned matrices'. Bulletin of the London Mathematical Society 2009 41(5):817-822

[18] Solymosi, Jozsef. Arithmetic Progressions in Sets with Small Sumsets'. Combinatorics, Probability and Computing. 15 (2006): 597 - 603.

[19] Solymosi, Jozsef; Tóth, Csaba D. Distinct distances in homogeneous sets in Euclidean space. Discrete Comput. Geom. 35 (2006), no. 4, 537–549.

[20] Solymosi, Jozsef. 'Regularity, uniformity, and quasirandomness'. Proceedings of the National Academy of Sciences of the United States of America. 102.23 (2005): 8075 – 8076.

[21] Solymosi, Jozsef. 'Note on a generalization of Roth's theorem'. Discrete and computational geometry; Algorithms Combin. Vol. 25. Ed. Janos Pach. Springer, 2003. 825 – 827.

[22] Chang, Mei-Chu and Jozsef Solymosi. 'Sum-product theorems and incidence geometry'. JOURNAL OF THE EUROPEAN MATHEMATICAL SOCIETY. 9.3 (2007): 545 - 560.

[23] Solymosi, Jozsef and Frank De Zeeuw. 'On a question of Erdos and Ulam'. Discrete and Computational Geometry, Volume 43, Issue 2 (2010), Page 393–401.

[24] Bourgain, J.; Katz, N.; Tao, T. A sum-product estimate in finite fields, and applications. Geom. Funct. Anal. 14 (2004), no. 1, 27–57.

[25] Hart,D. Iosevich, A. and Solymosi, J. Sum product estimates in finite fields via Kloosterman sums, INTERNATIONAL MATHEMATICS RESEARCH NOTICES. 2007: 1–14.

[26] Elekes, György, On the Number of Sums and Products, Acta Arithmetica LXXXI.4, (1997) 365-367

[27] Elekes, György; Nathanson, Melvyn B.; Ruzsa, Imre Z. Convexity and sumsets. J. Number Theory 83 (2000), no. 2, 194–201.

[28] Erdős, P. On some of my favourite theorems. Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993), 97–132, Bolyai Soc. Math. Stud., 2, János Bolyai Math. Soc., Budapest, 1996.

[29] W.T. Gowers, *Lower bounds of tower type for Szemerédi's Uniformity Lemma*, Geom. Funct. Anal 7, 1997, no. 2, pp. 322-337.

[30] J. Komlós, *The Blow-up Lemma*, Combinatorics, Probability and Computing, 8, 1999, pp. 161-176.

[31] J. Komlós, M. Simonovits, *Szemerédi's Regularity Lemma and its applications in graph theory*, in Combinatorics, Paul Erdős is Eighty (D. Miklós, V.T. Sós, and T. Szőnyi, Eds.), pp. 295-352, Bolyai Society Mathematical Studies, Vol. 2, János Bolyai Mathematical Society, Budapest, 1996.

[32] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arithmetica 27, 1975, pp. 199-245.

[33] E. Szemerédi, *Regular partitions of graphs*, Colloques Internationaux C.N.R.S. Nº 260 - Problèmes Combinatoires et Théorie des Graphes, Orsay, 1976, pp. 399-401.

[34] T. Tao, *A variant of the hypergraph removal lemma*, Journal of Combinatorial Theory, Ser. A 113, 2006, pp. 1257-1280.

[35] T. Tao, *Szemerédi's regularity lemma revisited*, Contrib. Discrete Math. 1, 2006, pp. 8-28.

dc_52_10

1. *Első fejezet:* A Hypergraph Removal Lemma alkalmazásai a Szemerédi tétel általánositásában.

    a. Bevezetés: “Regularity, uniformity, and quasirandomness”

    b. Első cikk: Roth tételének általánositása.

    c. Második cikk: Erdős és Graham egy problémájáról.

    d. Harmadik cikk: Számtani sorozatok kis összegű halmazokban

2. *Második fejezet:* Az összeg-szorzat problémáról

    e. Negyedik cikk: Összeg-szorzat becslés komplex számokra

    f. Ötödik cikk: Javitott becslés a Szemerédi-Trotter tétel alkalmazásával

    g. Hatodik cikk: Összeg-szorzat becslés mátrixokra

    h. Hetedik cikk: További javitás elemi lineáris algebra alkalmazásával

3. *Harmadik fejezet:* Számelméleti eredmények alkalmazása a diszkrét geometria területén.

    i. Nyolcadik cikk: Extremális pont-egyenes illeszkedési rendszerek lokális struktúrája

    j. Kilencedik cikk: Összeg-szorzat becslések geometriai alkalmazása

    k. Tizedik cikk: Erdős és Ulam egy problémájáról

4. *Negyedik fejezet:* Különböző távolságok

    l. Tizenegyedik cikk: A különböző távolságok száma magasabb dimenzióban

    m. Tizenkettedik cikk: Különböző távolságok homogén ponthalmazokban

dc_52_10

Regularity, uniformity, and quasirandomness

Jozsef Solymosi[†]

Department of Mathematics, University of British Columbia, 1984 Mathematics Road, Vancouver, BC, Canada V6T 1Z2

Graph theory is the appropriate language for discussing binary relations on objects. Results in graph theory have numerous applications in biology, chemistry, computer science, and physics. In cases of multiple relations, instead of binary relations more general structures known as hypergraphs are the right tools. However, it turns out that because of their extremely complex structure, hypergraphs are very difficult to deal with. As with number theory, there are questions about hypergraphs that are easy to state but very difficult to answer. In this issue of PNAS, Rödl *et al.* (1) extend a powerful tool, the regularity lemma, from graphs to hypergraphs.

Contrary to the general terminology, in extremal graph theory regularity is a measure of randomness. Random graphs are easy to work with, especially when one wants to estimate the (expected) number of small subgraphs. In complex structures, like in dense graphs, one can substitute randomness with weaker but still useful properties. The motivation behind graph regularity is to arrange the vertices of a graph in such a way that the graph becomes similar to the union of a few random graphs, and then one can apply standard counting methods from probability theory. In order to define hypergraph regularity, one has to introduce somehow complicated and technical notations. However, even without these notations we can formulate the most important consequence of the so-called hypergraph regularity method. The method, which is the combination of the hypergraph regularity lemma and a counting lemma is described by Rödl *et al.* (1). Similar results with the same consequences have been obtained independently by Gowers (2). Inspired by the methods of refs. 1 and 2, very recently Tao (T. Tao, personal communication) gave another proof of the main results. The road to the hypergraph regularity and counting lemmas was long and challenging.

Graph Regularity

Graph regularity was first introduced by Szemerédi (3), who used it to prove his celebrated theorem that every dense subset of integers contains arbitrary long arithmetic progressions. Today, one of the main tools in extremal graph theory is Szemerédi's regularity lemma (4), which makes arbitrary (usually large and dense) graphs manageable.[‡] It was widely expected that hypergraph regularity could provide a similarly useful tool to deal with hypergraphs. The problem is that one can easily formulate fake hypergraph regularity lemmas by simply generalizing the original regularity lemma. The question was if one can find the "right" hypergraph lemma that can be used to prove theorems that do not follow from an application of the ordinary regularity lemma. Chung (5) was the first to come up with generalizations of regularity; however, her result had certain limitations. Her findings were not strong enough for applications to Szemerédi-type theorems, but still they formed a significant precursor to the more modern hypergraph regularity lemmas. After several years of hard work, Rödl and his students (1) have devised a solution providing a right notation of hypergraph regularity and proving the corresponding theorems using purely combinatorial tools. Gowers' approach (2) is somehow different, more analytic. The notations and proofs are related to his earlier proof of Szemerédi's theorem using Fourier analysis (6). One should mention here that the Cauchy–Schwarz-type arguments Gowers uses in his counting lemma were very influential in the recent results of Green and Tao (7) on long arithmetic progressions in the primes.

Rödl *et al.* extend a powerful tool, the regularity lemma, from graphs to hypergraphs.

An Important Corollary

Graphs and hypergraphs are general combinatorial objects. A graph G is given by its vertex set V (G) and the edge set $E(G)$, a list of vertex pairs that are connected by an edge. The notation of a hypergraph is similar. Given a set S as the vertex set, a family of the subsets of S will define the hyperedges. In this paper, we will focus on k-uniform hypergraphs, on hypergraphs where all the edges have the same size, k. With this notation, the two uniform hypergraphs are the ordinary graphs.

Given a k-uniform hypergraph, H_k^n, on an n-element vertex set, $V(H_k^n)$ a clique, K_{k+1}, is a $k+1$-element subset of $V(H_k^n)$ such that any k-tuple of K_{k+1} is an edge of the hypergraph H_k^n. Two cliques are said to be edge-disjoint if they don't have a common edge. Any set of pairwise edge-disjoint cliques in H_k^n has cardinality at most $\binom{n}{k}/(k+1)$ because every clique has $k+1$ edges. The main result of ref. 1 is that if a hypergraph contains a large set, S, of pairwise edge-disjoint cliques, then it contains many cliques. In particular, the hypergraph contains at least one clique that is not in S. We will refer to the result below as the *Removal Lemma* for k-uniform hypergraphs. The reason why it is called *Removal Lemma* is that one can formulate the statement in the following equivalent way. If a hypergraph contains few cliques, then after removing only few edges from the hypergraph, the remaining hypergraph will not contain cliques at all.

Removal Lemma. *For any $c > 0$ real number and $k \geq 2$ integer, there is a $\delta > 0$ that depends on c and k only, such that the following is true. If H_k^n contains a set, S, of pairwise edge-disjoint cliques with cardinality $|S| \geq c\binom{n}{k}$, then H_k^n contains at least $\delta\binom{n}{k+1}$ cliques.*

A typical application of the result would be as follows. We want to prove that a given hypergraph contains two cliques sharing an edge. If we can show that there is a large set of pairwise edge-disjoint cliques, then we are done. To illustrate the method, we prove a generalization of Roth's theorem (8) about three-term arithmetic progressions in dense subsets of integers. We will show that if S is a dense subset of a large $N \times N$ integer grid, then S contains an isosceles equilateral triangle, three points with coordinates (x, y),

See companion article on page 8109.

[†]E-mail: solymosi@math.ubc.ca.

[‡]For a graph $G = (V, E)$ and two disjoint sets $V_1, V_2 \subset V$, we denote by $E(V_1, V_2)$ the set of edges with one endpoint in V_1 and one endpoint in V_2. The density $d(V_1, V_2)$ is given by $d(V_1, V_2) = |E(V_1, V_2)|/(|V_1||V_2|)$. We say that the graph induced by V_1, V_2 is ε-regular if for all $V_1^* \subset V_1$ and $V_2^* \subset V_2$ with $|V_1^*| \geq \varepsilon|V_1|$ and $|V_2^*| \geq \varepsilon|V_2|$, $|d(V_1^*, V_2^*) - d(V_1, V_2)| \leq \varepsilon$. Szemerédi's regularity lemma claims that for any $\varepsilon > 0$ there is a number, $t = t(\varepsilon)$, such that any graph's vertex set can be partitioned into t almost equal vertex classes such that with only εt^2 exemptions the bipartite graphs between the classes are ε-regular.

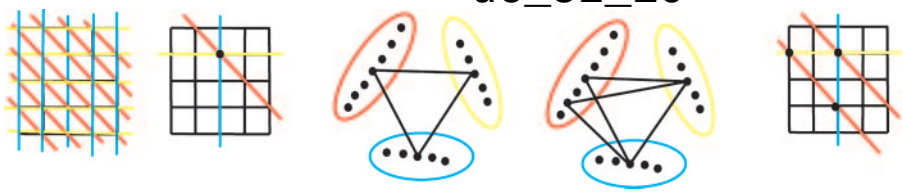


Fig. 1. Take a tripartite graph in which the vertices of the graph are the red, yellow, and green lines and the edges are defined by the set *S*. Two vertices are connected by an edge if the crossing point of the corresponding lines is a point of *S*. A triangle in the graph corresponds to three lines such that any two intersect in a point of *S*. If there are two triangles sharing an edge, then at least one triangle is not degenerate; thus, we have an isosceles equilateral triangle in *S*. If *S* is a dense subset of a large grid, then by the *Triangle Removal Lemma* there are many triangles in the graph. Therefore, there is an edge that is the edge of two triangles, so *S* contains an isosceles equilateral triangle.

$(x + d, y)$, and $(x, y + d)$, where d is a non-zero integer. It is easy to see that the statement implies Roth's theorem (Fig. 1).

The very same trick can be applied for higher dimensional grids, hyperplanes, and hypergraphs. This calculation leads us to a combinatorial proof of the so-called multidimensional Szemerédi theorem, which was proved by Fürstenberg and Katznelson (9) using ergodic theory.

It is not known how δ depends on c. Even in the simplest case, $k = 2$, the gap between the best known upper and lower bounds is huge. When n is large enough, $\delta\binom{n}{k+1}$ is larger than $\binom{n}{k}/(k+1)$, so there is at least one clique in H_k^n that is not in S. It is surprising that this seemingly weak statement needs such heavy machinery. In most of the applications, all we need is to show that in a hypergraph there are two cliques that have a common edge. Random hypergraphs almost surely have such a pair of cliques. Therefore, if one can show that a given hypergraph is somehow similar to the random hypergraph, then this could lead to the proof. What we want from a hypergraph regularity lemma is to find for a given hypergraph, H_k^n a partition of the one-, two-, three-, . . . , $(k - 1)$-element subsets of $V(H_k^n)$ into few classes such that the subgraphs, spanned by the classes, behave in a random-like way with only few exceptions. Also, one should come up with the right definition of "random-like." This plan is nice, but unfortunately for $k > 2$ the solution is quite complicated. In 1978, for $k = 2$, Ruzsa and Szemerédi (10) proved that graph regularity implies the *Removal Lemma* for graphs. What Ruzsa and Szemerédi proved by using the regularity lemma for graphs is the following.

Triangle Removal Lemma. *If a graph on n vertices contains at least cn^2 edge disjoint triangles, then it contains at least δn^3 triangles.*

It was 25 years later when Frankl and Rödl (11) published the $k = 3$ case. This shows how difficult it was to find the right generalization of graph regularity to hypergraphs. There is a test to decide whether a hypergraph regularity is useful or not. Does it imply the *Removal Lemma*? If the answer is yes, then it is a correct concept of regularity indeed. On the contrary, applications of the hypergraph regularity could go beyond the *Removal Lemma*. There are already examples for which the hypergraph regularity method, combined with ergodic theory, analysis, and number theory, are used efficiently to solve difficult problems in mathematics.

1. Rödl, V., Nagle, B., Skokan, J., Schacht, M. & Kohayakawa, Y. (2005) *Proc. Natl. Acad. Sci. USA* **102,** 8109–8113.
2. Gowers, W. T. (2005) *Comb. Probab. Comput*., in press.
3. Szemerédi, E. (1975) *Acta Arith.* **27,** 199–245.
4. Szemerédi, E. (1978) in *Problèmes Combinatoires et Théorie des Graphes* (Centre Natl. Rech. Sci., Paris), pp. 399–401.
5. Chung, F. R. K. (1990) *Random Struct. Algorithms* **1,** 363–382.
6. Gowers, W. T. (2001) *Geom. Funct. Anal.* **11,** 465–588.
7. Green, B. & Tao, T. (2005) *Ann. Math.,* in press.
8. Roth, K. F. (1953) *J. London Math. Soc.* **28,** 104–109.
9. Fürstenberg, H. & Katznelson, Y. (1978) *J. Anal. Math.* **34,** 275–291.
10. Ruzsa, I. & Szemerédi, E. (1976) *Comb. Coll. Math. Soc. J. Bolyai* **18,** 939–945.
11. Frankl, P. & V. Rödl, V. (2002) *Random Struct. Algorithms* **20,** 131–164.

Note on a Generalization of Roth's Theorem

József Solymosi

Abstract

We give a simple proof that for sufficiently large N, every subset of $[N]^2$ of size at least δN^2 contains three points of the form $\{(a,b),(a+d,b),(a,b+d)\}$.

In this note we give a simple proof for a theorem of Ajtai and Szemerédi [1]. In their proof Ajtai and Szemerédi used and iterated Szemerédi's theorem about long arithmetic progressions in dense sets of integers [8]. A more general theorem of Fürstenberg and Katznelson also implies Theorem 1, but does not give bound on N as it uses ergodic theory [2]. After improving the bound in Szemerédi's theorem, Gowers asked for a quantitative proof of Theorem 1 [3, 4].

Theorem 1 *(Ajtai-Szemerédi) For any real number $\delta > 0$ there is a natural number N_0 such that for $N > N_0$ every subset of $[N]^2$ of size at least δN^2 contains a triple of the form $\{(a,b),(a+d,b),(a,b+d)\}$ for some integer $d \neq 0$.*

The key of the proof is a lemma of Ruzsa and Szemerédi [7]. A subgraph of a graph G is a *matching* if every vertex has degree one. A matching M is an *induced matching* if there are no other edges of G between the vertices of M.

Lemma 2 *(Ruzsa-Szemerédi) If G_n is the union of n induced matchings, then $e(G_n) = o(n^2)$.*

The lemma, with a simple proof deduced from Szemerédi's Regularity Lemma, can be also found in a survey paper of Komlós and Simonovits [5].

Proof of Theorem 1: Let S be a subset of the grid $[N]^2$ of size at least δN^2. We refer to a point of the grid with its coordinates, which are pairs $(i,j); i,j \in \{1,2,\dots,N\}$. Let us define a bipartite graph $G(A,B)$ with vertex sets $A = \{v_1,\dots,v_N\}$ and $B = \{w_1,\dots,w_N\}$. Two vertices v_i and w_j are connected by an edge iff $(i,j) \in S$ (see Fig. 1).

Let us partition the edges of G according to their length, $(v_i,w_j) \sim (v_l,w_m)$ iff $i+j = l+m$. Every partition class is a matching, so we can apply Lemma 2 to G. If N is large enough, then at least one matching is not induced. A triple

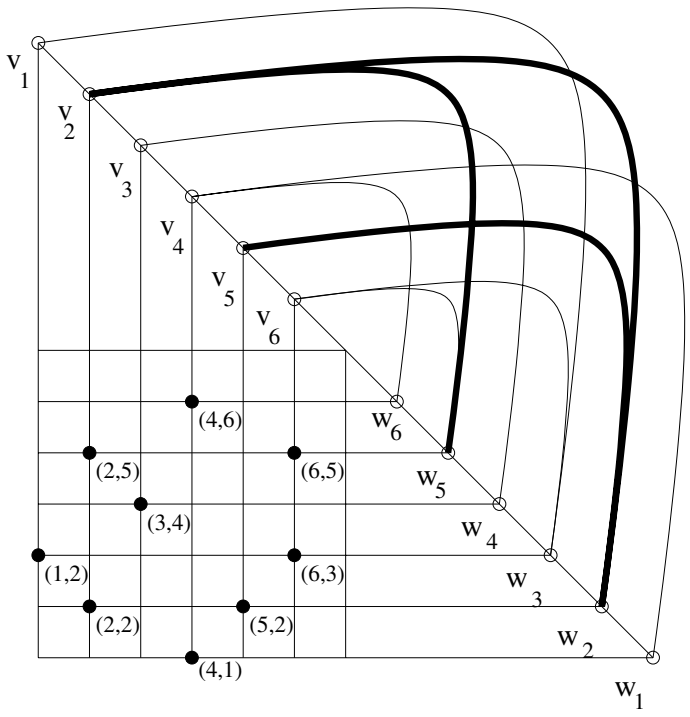


Figure 1: Converting points into edges

of edges $(v_i, w_m), (v_i, w_j), (v_l, w_m)$ such that $(v_i, w_j) \sim (v_l, w_m)$ guarantees a triple in S, $\{(a, b), (a + d, b), (a, b + d)\}$ (see bold edges in Fig.1). □

The only known proof of Lemma 2 uses Szemerédi's Regularity Lemma [8], so while the proof is quantitative, it gives a tower-type bound on $N_0 = N_0(\delta^{-1})$. It would be very important to find another, maybe analytical proof for Lemma 2 to get a better bound.

References

[1] M. Ajtai and E. Szemerédi, Sets of Lattice Points That Form No Squares. *Studia Scientiarum Mathematicarum Hungarica.* **9** (1974), 9–11.

[2] H. Fürstenberg and Y. Katznelson, A density version of the Hales-Jewett theorem. *J.d'Analyse Math.***57** (1991), 64–119.

[3] W.T. Gowers, A new proof of Szemerédi's theorem. *GAFA, Geom. Funct. Anal.* **11** (2001) 465–588.

[4] W.T. Gowers, Rough structure and classification. *GAFA, Geom. Funct. Anal.* Special Volume - GAFA2000 "Visions in Mathematics", Tel Aviv, 1999. Part I, 79–117.

[5] J. Komlós and M. Simonovits, Szemerédi's Regularity Lemma and its applications in graph theory. in: Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993), 295–352, Bolyai Soc. Math. Stud., 2, János Bolyai Math. Soc., Budapest, 1996.

[6] K.F. Roth On certain sets of integers, *J.London Math. Soc.***28** (1953), 245–252.

[7] I.Z. Ruzsa and E. Szemerédi, Triple systems with no six points carrying three triangles. in: Colloquia Mathematica Societatis János Bolyai, 18. Combinatorics, Keszthely (Hungary), 1976, 939–945.

[8] E. Szemerédi, On sets of integers containing no k elements in arithmetical progression. *Acta Arithmetica* **27** (1975), 199–245.

About Author

József Solymosi is at the Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112, USA. *solymosi@math.ucsd.edu.*

Acknowledgments

I thank Timothy Gowers and Tibor Szabó for useful discussions. Work on this paper has been supported by the Berlin-Zürich European Graduate Program "Combinatorics, Geometry, and Computation" and by the Computer and Automation Research Institute, Hungarian Academy of Sciences.

Combinatorics, Probability and Computing (2004) 13, 263–267. © 2004 Cambridge University Press
DOI: 10.1017/S0963548303005959 Printed in the United Kingdom

A Note on a Question of Erdős and Graham

J. SOLYMOSI†

Department of Mathematics, University of California in San Diego,
9500 Gilman Drive, La Jolla CA 92093-0112, USA
(e-mail: solymosi@math.ucsd.edu)

Received 29 August 2002; revised 17 November 2002

We give a quantitative proof that, for sufficiently large N, every subset of $[N]^2$ of size at least δN^2 contains a square, *i.e.*, four points with coordinates $\{(a,b),(a+d,b),(a,b+d),(a+d,b+d)\}$.

1. Introduction

In this note we prove a generalization of Szemerédi's theorem about arithmetic progressions of length four [12]. This generalization, Theorem 1.1, was first considered by Ron Graham in 1970 and conjectured by him and Erdős (published in [2] and [1]). Using Szemerédi's deep theorem [11] about arithmetic progressions of length k, Ajtai and Szemerédi [1] proved a simpler statement: for sufficiently large N, every subset of $[N]^2$ of size at least δN^2 contains three points with coordinates $\{(a,b),(a+d,b),(a,b+d)\}$. ($[N]=\{0,1,2,\ldots,N-1\}$). Later Fürstenberg and Katznelson proved a much stronger general theorem [3] (see Theorem 3.1), but their proof does not give an explicit bound as it uses ergodic theory. After giving an analytic proof for Szemerédi's theorem, Tim Gowers again raised the question of finding a quantitative proof for Graham's question [5, 6]. Using a recent result of Frankl and Rödl we give a combinatorial proof for this theorem.

Theorem 1.1. *For any real number $\delta > 0$ there is a natural number $N_0 = N_0(\delta)$ such that for $N > N_0$ every subset of $[N]^2$ of size at least δN^2 contains a square, i.e., a quadruple of the form $\{(a,b),(a+d,b),(a,b+d),(a+d,b+d)\}$ for some integer $d \neq 0$.*

† Supported by the Berlin–Zürich European Graduate Program 'Combinatorics, Geometry, and Computation' and by MTA SZTAKI. Present address: Department of Mathematics, University of British Columbia, BC, Vancouver V6T 1Y4, Canada (e-mail: solymosi@math.ubc.ca).

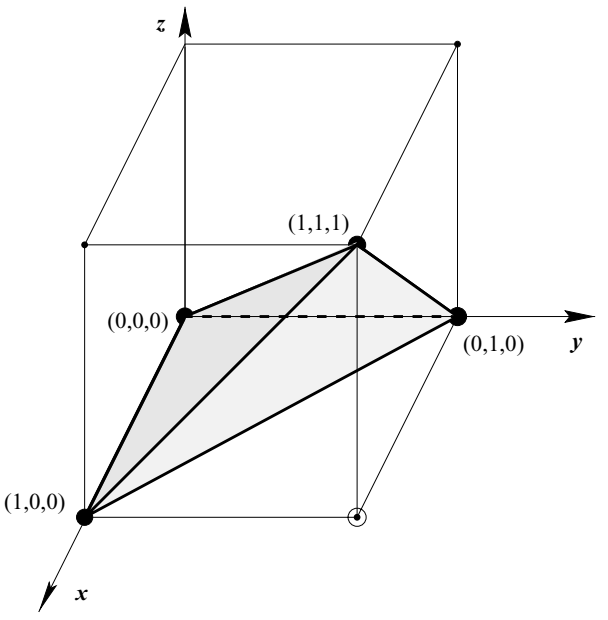


Figure 1. A quadruple of the form (1.1)

Before Theorem 1.1 we prove the following theorem.

Theorem 1.2. *For any real number $\delta > 0$ there is a natural number $N_0 = N_0(\delta)$ such that, for $N > N_0$, every subset of $[N]^3$ of size at least δN^3 contains a quadruple of the form*

$$\{(a,b,c),(a+d,b,c),(a,b+d,c),(a+d,b+d,c+d)\} \tag{1.1}$$

for some integer $d \neq 0$.

Proposition 1.3. *Theorem* 1.2 *implies Theorem* 1.1. □

Proof. Let us suppose that Theorem 1.1 is false. Then there is a real number $\delta > 0$ and, for every N, a subset S_N of $[N]^2$, such that $|S_N| > \delta N^2$ and S_N does not contain any square. For every S_N we can define a subset of $[N]^3$ by lifting up all the points of S_N into 3D:

$$S_N^* = \{(a,b,c) : (a,b) \in S_N, c \in [N]\}.$$

The size of S_N^* is larger than δN^3 and does not contain any quadruple of the form (1.1). This contradicts Theorem 1.2. □

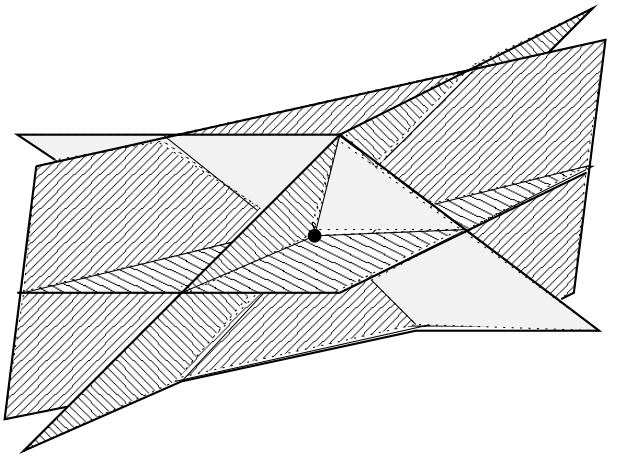


Figure 2. Every point of S defines $\binom{4}{3}$ edges in $\mathscr{H}$

2. Proof

Proof of Theorem 1.2. We define a three-uniform hypergraph $\mathscr{H}$. The vertex set $V(\mathscr{H})$ is a collection of planes:

$$\begin{aligned} a_i &= \{z = i\} && \text{and } V_1 = \{a_i : 0 \leqslant i \leqslant N-1\},\\ b_i &= \{-x + z = i\} && \text{and } V_2 = \{b_i : -N+1 \leqslant i \leqslant N-1\},\\ c_i &= \{-y + z = i\} && \text{and } V_3 = \{c_i : -N+1 \leqslant i \leqslant N-1\},\\ d_i &= \{x + y - z = i\} && \text{and } V_4 = \{d_i : -N+1 \leqslant i \leqslant 2N-2\},\\ V(\mathscr{H}) &= V_1 \cup V_2 \cup V_3 \cup V_4. \end{aligned}$$

These are the planes parallel with the faces of any simplex given by (1.1) and have points from $[N]^3$. The edge set $E(\mathscr{H})$ is defined by a point set $S \subset [N]^3$. Three distinct vertices v_1, v_2, and v_3 form an edge if the intersection point of the corresponding planes p_1, p_2 and p_3 is in S, that is,

$$E(\mathscr{H}) = \{(v_1, v_2, v_3) : v_i \in V (1 \leqslant i \leqslant 3), p_1 \cap p_2 \cap p_3 \in S\}.$$

$\mathscr{H}$ is a 4-partite hypergraph with classes V_1, V_2, V_3, and V_4. We are going to show that if S does not contain any quadruple like (1.1), then $|E(\mathscr{H})|$ – and therefore also $|S|$ – is $o(N^3)$. This will prove Theorem 1.2.

The next conjecture is a special case of a more general conjecture of Frankl and Rödl [8]. A subgraph in a k-uniform hypergraph is a *complete subgraph* if it has at least $k+1$ vertices and all k-tuples of its vertices are edges.

Conjecture 2.1. *Given an integer $k \geqslant 2$. If $\mathscr{G}$ is a k-uniform hypergraph such that every edge is an edge of exactly one complete subgraph, then the number of edges $|E(\mathscr{G})|$ is $o(|V(\mathscr{G})|^k)$.*

For $k = 2$ the conjecture is equivalent to the so-called (6,3)-theorem proved by Ruzsa and Szemerédi [10], and the $k = 3$ case was proved by Frankl and Rödl [8].

Theorem 2.2. (Frankl and Rödl) *If $\mathcal{G}$ is a 3-uniform hypergraph such that every edge is an edge of exactly one complete subgraph, then the number of edges $|E(\mathcal{G})|$ is $o(|V(\mathcal{G})|^3)$.*

Remark. In their proof Frankl and Rödl applied Szemerédi's Regularity Lemma; therefore here we cannot achieve more than a tower-type upper bound on N_0 in Theorem 1.1. (For the details of why, in general, the Regularity Lemma gives only a weak bound, we refer to the paper of Gowers [4].)

In $\mathcal{H}$ four vertices a_i, b_j, c_k, and d_l form a complete subgraph if any triple has its intersection point in S. If the planes are not concurrent planes, *i.e.*, $a_i \cap b_j \cap c_k \cap d_l = \emptyset$, then a_i, b_j, c_k, and d_l is a quadruple like (1.1), *i.e.*, the intersection points of the triples form a simplex similar to $\{(0,0,0),(1,0,0),(0,1,0),(1,1,1)\}$, because the corresponding faces are parallel. Let us suppose that there is no such quadruple in S. Then every edge of $\mathcal{H}$ is an edge of exactly one complete subgraph, and $|E(\mathcal{H})| = o(N^3)$ by Theorem 2.2. □

3. Conjectures

If Conjecture 2.1 was true, then it would imply the following 'multidimensional Szemerédi theorem' [3].

Theorem 3.1. (Fürstenberg and Katznelson) *For any real number $\delta > 0$ and positive integers K, d there is a natural number $N_0 = N_0(\delta, K, d)$ such that for $N > N_0$ every subset of $[N]^d$ of size at least δN^d contains a homothetic copy of $[K]^d$.*

We state a special case of Conjecture 2.1. It would also imply Theorem 3.1 following the steps of the proof of Theorem 1.1 in higher dimensions, and as a plus there is some geometry which could be useful for a possible proof.

Conjecture 3.2. *For any real number $\delta > 0$ and positive integer d there is a natural number $N_0 = N_0(\delta, d)$ such that, for $N > N_0$, any set of N hyperplanes S and at least δN^d points, where every point is an element of at least $d+1$ hyperplanes, contains a simplex (i.e., $d+1$ distinct points such that any d-tuples are contained by a hyperplane from S).*

We close this note with a nice conjecture of Graham [7] which, if true, would give a sufficient condition for the existence of a square in an infinite lattice set.

Conjecture 3.3. (Graham) *Given a set of lattice points in the plane*

$$S = \{p_1, p_2, \ldots, p_i, p_{i+1}, \ldots\},$$

let us denote the distance of p_i from the origin by d_i. If

$$\sum_{i=1}^{\infty} \frac{1}{d_i^2} = \infty,$$

then S contains the four vertices of an axes-parallel square.

Acknowledgement

I thank Ron Graham and Tim Gowers for useful discussions and for their help.

References

[1] Ajtai, M. and Szemerédi, E. (1974) Sets of lattice points that form no squares. *Studia Scientiarium Mathematicarum Hungarica* **9** 9–11.
[2] Erdős, P. (1973) Problems and results on combinatorial number theory. In *A Survey of Combinatorial Theory* (*Proc. Internat. Sympos., Colorado State Univ., Fort Collins, Colo., 1971*), North-Holland, Amsterdam, pp. 117–138.
[3] Fürstenberg, H. and Katznelson, Y. (1991) A density version of the Hales–Jewett theorem. *J. d'Analyse Math.* **57** 64–119.
[4] Gowers, W. T. (1997) Lower bounds of tower type for Szemeredi's uniformity lemma. *GAFA* (*Geometric and Functional Analysis*) **7** 322–337.
[5] Gowers, W. T. (2001) A new proof of Szemerédi's theorem. *GAFA* (*Geometric and Functional Analysis*) **11** 465–588.
[6] Gowers, W. T. (1999) Rough structure and classification. *GAFA* (*Geometric and Functional Analysis*), *Special Volume: GAFA2000 'Visions in Mathematics', Tel Aviv*, Part I, 79–117.
[7] Graham, R. (1997) Conjecture 8.4.6. In *Discrete and Computational Geometry* (J. E. Goodman and J. O'Rourke, eds), CRC Press, Boca Raton, NY, p. 11.
[8] Frankl, P. and Rödl, V. (2002) Extremal problems on set systems. *Random Struct. Alg.* **20** 131–164.
[9] Roth, K. F. (1953) On certain sets of integers. *J. London Math. Soc.* **28** 245–252.
[10] Ruzsa, I. Z. and Szemerédi, E. (1976) Triple systems with no six points carrying three triangles. In *Combinatorics, Keszthely* (*Hungary*), Vol. 18 of *Colloquia Mathematica Societatis János Bolyai*, pp. 939–945.
[11] Szemerédi, E. (1975) On sets of integers containing no k elements in arithmetic progression. *Acta Arithmetica* **27** 199–245.
[12] Szemerédi, E. (1969) On sets of integers containing no four elements in arithmetic progression. *Acta Math. Acad. Sci. Hungar.* **20** 89–104. Also in *Number Theory* (*Colloq., János Bolyai Math. Soc., Debrecen, 1968*), North-Holland, Amsterdam, pp. 197–204.

Combinatorics, Probability and Computing (2006) 15, 597–603. © 2006 Cambridge University Press
doi:10.1017/S0963548306007516 Printed in the United Kingdom

Arithmetic Progressions in Sets with Small Sumsets

JÓZSEF SOLYMOSI†

Department of Mathematics, University of British Columbia,
1984 Mathematics Road, Vancouver, BC, Canada V6T 1Z2
(e-mail: solymosi@math.ubc.ca)

Received 11 January 2004; revised 24 August 2004

We present an elementary proof that if A is a finite set of numbers, and the sumset $A +_G A$ is small, $|A +_G A| \leqslant c|A|$, along a dense graph G, then A contains k-term arithmetic progressions.

1. Introduction

A well-known theorem of Szemerédi [15] states that every dense subset of integers contains long arithmetic progressions. A different, but somehow related result of Freiman [5] says that if the sumset of a finite set of numbers A is small, *i.e.*, $|A + A| \leqslant C|A|$, then A is the subset of a (not very large) generalized arithmetic progression. Balog and Szemerédi proved in [1] that a similar structural statement holds under weaker assumptions. (For correct statements and details, see [8].) As a corollary of their result, Freiman's theorem, and Szemerédi's theorem about k-term arithmetic progressions, Balog and Szemerédi proved Theorem 1.1 below. The goal of this paper is to present a simple, purely combinatorial proof of this assertion.

Let A be a set of numbers and G be a graph such that the vertex set of G is A. The *sumset of A along G* is

$$A +_G A = \{a + b : a, b \in A \text{ and } (a, b) \in E(G)\}.$$

Theorem 1.1. *For every $c, K, k > 0$ there is a threshold $n_0 = n_0(c, K, k)$ such that if $|A| = n \geqslant n_0$, $|A +_G A| \leqslant K|A|$, and $|E(G)| \geqslant cn^2$, then A contains a k-term arithmetic progression.*

† Research partially supported by NSERC and OTKA grants.

2. Lines and hyperplanes

There are arrangements of n lines on the Euclidean plane such that the maximum number of points incident with at least three lines is $\frac{n^2}{6}$. Not much is known about the structure of arrangements where the number of such points is close to the maximum, say cn^2, where c is a positive constant. Nevertheless, the following is true.

Lemma 2.1. *For every $c > 0$ there is a threshold $n_0 = n_0(c)$ and a positive $\delta = \delta(c)$ such that, for any set of $n \geqslant n_0$ lines L and any set of $m \geqslant cn^2$ points P, if every point is incident to three lines, then there are at least δn^3 triangles in the arrangement.* (*A triangle is a set of three distinct points from P such that any two are incident to a line from L.*)

Proof. This lemma is implied by the following theorem of Ruzsa and Szemerédi [13].

Theorem 2.2. ([13]) *Let G be a graph on n vertices. If G is the union of cn^2 edge-disjoint triangles, then G contains at least δn^3 triangles, where δ depends on c only.*

To prove Lemma 2.1, let us construct a graph where L is the vertex set, and two vertices are adjacent if and only if the corresponding lines cross at a point of P. This graph is the union of cn^2 disjoint triangles, and every point of P defines a unique triangle, so we can apply Theorem 2.2. □

The result above suffices to prove Theorem 1.1 for 3-term arithmetic progressions. But for larger values of k, we need a generalization of Lemma 2.1.

Lemma 2.3. *For every $c > 0$ and $d \geqslant 2$, there is a threshold $n_0 = n_0(c, d)$ and a positive $\delta = \delta(c, d)$ such that, for any set of $n \geqslant n_0$ hyperplanes L and any set of $m \geqslant cn^d$ points P, if every point is incident to $d + 1$ hyperplanes, then there are at least δn^{d+1} simplices in the arrangement.* (*A simplex is a set of $d + 1$ distinct points from P such that any d are incident to a hyperplane from L.*)

Lemma 2.3 follows from the Frankl–Rödl conjecture [4], the generalization of Theorem 2.2. The $d = 3$ case was proved in [4] and the conjecture has been proved recently by Gowers [6] and independently by Nagle, Rödl, Schacht and Skokan [7, 10]. For details on how Lemma 2.3 follows from the Frankl–Rödl conjecture, see [14].

3. The $k = 3$ case

Let A be a set of numbers and G be a graph such that the vertex set of G is A. We define the *difference-set of A along G* as

$$A -_G A = \{a - b : a, b \in A \text{ and } (a, b) \in E(G)\}.$$

Lemma 3.1. *For every $\epsilon, c, K > 0$ there is a number $D = D(\epsilon, c, K)$ such that, if $|A +_G A| \leqslant K|A|$ and $|E(G)| \geqslant c|A|^2$, then there is a graph $G' \subset G$ such that $|E(G')| \geqslant (1 - \epsilon)|E(G)|$ and $|A -_{G'} A| \leqslant D|A|$.*

Proof. Let us consider the arrangement of points given by a subset of the Cartesian product $A \times A$ and the lines $y = a$, $x = a$ for every $a \in A$, and $x + y = t$ for every $t \in A +_G A$. The pointset P is defined by $(a, b) \in P$ if and only if $(a, b) \in E(G)$. By Lemma 2.1, the number of triangles in this arrangement is δn^3. The triangles here are right isosceles triangles. We say that a point in P is *popular* if the point is the right-angle vertex of at least αn triangles. Selecting $\alpha = \frac{\delta(\epsilon c)}{\epsilon c}$, where $\delta(\cdot)$ is the function from Lemma 2.1, all but at most $\epsilon c n^2$ points of P are popular.

A $t \in A - A$ is *popular* if $|\{(a, b) : a - b = t; a, b \in A\}| \geqslant \alpha n$. The number of popular ts is at most Dn, where D depends on α only. $A \times A$ is a Cartesian product; therefore every triangle can be extended to a square adding one extra point from $A \times A$. Every popular point p is the right-angle vertex of at least αn triangles. Therefore p is incident to a line $x - y = t$, where t is popular, because this line contains at least αn 'fourth' vertices of squares with p. □

Proof of Theorem 1.1, case $k = 3$. Let us apply Lemma 2.1 to the pointset P' defined by $(a, b) \in P'$ if and only if $(a, b) \in E(G')$ and the lines are $y = a$ for every $a \in A$, $x - y = t$ for every $t \in A -_{G'} A$, and $x + y = s$ for every $s \in A +_G A$. By Theorem 2.2, if $|A|$ is large enough, then there are triangles in the arrangement. The vertices of such triangles are vertices from $P' \subset A \times A$. The vertical lines through the vertices form a 3-term arithmetic progression and therefore A contains δn^2 3-term arithmetic progressions, where $\delta > 0$ depends on c only. □

4. The general, $k > 3$, case

Following the steps of the proof for $k = 3$, we prove the general case by induction on k. We prove the following theorem, which was conjectured by Erdős, and proved by Balog and Szemerédi in [1]. Theorem 4.1, together with the $k = 3$ case, gives a proof of Theorem 1.1.

Theorem 4.1. *For every $c > 0$ and $k > 3$ there is an n_0 such that, if A contains at least $c|A|^2$ 3-term arithmetic progressions and $|A| \geqslant n_0$, then A contains a k-term arithmetic progression.*

Instead of triangles, we must consider simplices. Set $k = d$. In the d-dimensional space we show that $A \times \cdots \times A$, the d-fold Cartesian product of A, contains a simplex in which the vertices' first coordinates form a $(d + 1)$-term arithmetic progression.

The simplices we are looking for are homothetic[1] images of the simplex S_d whose vertices are listed below:

$$
\begin{gathered}
(0, 0, 0, 0, \ldots, 0, 0) \\
(1, 1, 0, 0, \ldots, 0, 0) \\
(2, 0, 1, 0, \ldots, 0, 0) \\
(3, 0, 0, 1, \ldots, 0, 0) \\
\vdots \\
(d - 1, 0, \ldots, 1, 0) \\
(d, 0, 0, 0, \ldots, 0, 0).
\end{gathered}
$$

[1] Here we say that two simplices are homothetic if the corresponding facets are parallel.

An important property of S_d is that its facets can be pushed into a 'shorter' grid. The facets of S_d are parallel to hyperplanes, defined by the origin $(0,0,0,0,\dots,0,0)$, and some $(d-1)$-tuples of the grid

$$\{0,1,2,\dots,d-1\}\times\{-1,0,1\}\times\{0,1\}^{d-2}.$$

For example, if $d=3$, then the facets are

$$\{(0,0,0),(1,1,0),(2,0,1)\}$$
$$\{(0,0,0),(1,1,0),(3,0,0)\}$$
$$\{(0,0,0),(2,0,1),(3,0,0)\}$$
$$\{(1,1,0),(2,0,1),(3,0,0)\},$$

and the corresponding parallel planes in

$$\{0,1,2\}\times\{-1,0,1\}\times\{0,1\}$$

are the planes incident to the triples

$$\{(0,0,0),(1,1,0),(2,0,1)\}$$
$$\{(0,0,0),(1,1,0),(2,0,0)\}$$
$$\{(0,0,0),(2,0,1),(2,0,0)\}$$
$$\{(0,0,0),(1,-1,1),(2,-1,0)\}.$$

In general, if a facet of S_d contains the origin and the 'last point' $(d,0,0,0,\dots,0,0)$, then if we replace the later one by $(d-1,0,0,0,\dots,0,0)$, the new d-tuples define the same hyperplane. The remaining facet f, given by

$$(1,1,0,0,\dots,0,0)$$
$$(2,0,1,0,\dots,0,0)$$
$$(3,0,0,1,\dots,0,0)$$
$$\vdots$$
$$(d-1,0,\dots,1,0)$$
$$(d,0,0,0,\dots,0,0),$$

is parallel to the hyperplane through the vertices of $f-(1,1,0,0,\dots,0,0)$,

$$(0,0,0,0,\dots,0,0)$$
$$(1,-1,1,0,\dots,0,0)$$
$$(2,-1,0,1,\dots,0,0)$$
$$\vdots$$
$$(d-2,-1,\dots,1,0)$$
$$(d-1,-1,0,0,\dots,0,0).$$

In a homothetic copy of the grid

$$T_d=\{0,1,2,\dots,d-1\}\times\{-1,0,1\}\times\{0,1\}^{d-2},$$

the image of the origin is called the *holder* of the grid.

As the induction hypothesis, let us suppose that Theorem 4.1 is true for a $k \geqslant 3$ in a stronger form, provided that the number of k-term arithmetic progressions in A is at least $c|A|^2$.

Then the number of distinct homothetic copies of T_d in

$$\mathbb{A}_d = \underbrace{A \times \ldots \times A}_{d}$$

is at least $c'|A|^{d+1}$ (c' depends on c only). Let us say that a point $p \in \mathbb{A}_d$ is *popular* if p is the holder of at least $\alpha|A|$ grids. If p is popular, then, for any facet of S_d, f, the point p is the element of at least $\alpha|A|$ d-tuples, similar and parallel to f. If α is small enough, then at least $\gamma|A|^d$ points of $\mathbb{A}_d$ are popular, where γ depends on c and α only.

A hyperplane H is *β-rich* if it is incident to many points, $|H \cap \mathbb{A}_d| \geqslant \beta|A|^{d-1}$. For every facet of S_d, f, let us denote the set of β-rich hyperplanes which are parallel to f by $\mathscr{H}_f$.

Lemma 4.2. *For some choice of β, at least half of the popular points are incident to $d+1$ β-rich hyperplanes, parallel to the facets of S_d.*

Suppose to the contrary that, for a facet f, more than $\frac{\gamma}{2d}|A|^d$ popular points are not incident to hyperplanes of $\mathscr{H}_f$. Then, more than

$$\alpha|A|\frac{\gamma}{2d}|A|^d = \frac{\gamma\alpha}{2d}|A|^{d+1} \tag{4.1}$$

d-tuples, similar and parallel to f, are not covered by $\mathscr{H}_f$. Let us denote the hyperplanes incident to the 'uncovered' d-tuples by $L_1, L_2, \ldots, L_m$, and the number of points on the hyperplanes by $\mathscr{L}_1, \mathscr{L}_2, \ldots, \mathscr{L}_m$. A simple result of Elekes and Erdős [2, 3] implies that hyperplanes with few points cannot cover many d-tuples.

Theorem 4.3. ([3]) *The number of homothetic copies of f in L_i is at most $c_d \mathscr{L}_i^{1+1/(d-1)}$, where c_d depends on d only.*

The inequalities

$$\sum_{i=1}^{m} \mathscr{L}_i \leqslant |A|^d, \text{ and } \mathscr{L}_i \leqslant \beta|A|^{d-1}.$$

lead us to the proof of Lemma 4.2.

The number of d-tuples covered by L_is is at most

$$c_d \sum_{i=1}^{m} \mathscr{L}_i^{1+1/(d-1)} \leqslant c_d \frac{|A|^d}{\beta|A|^{d-1}}(\beta|A|^{d-1})^{1+1/(d-1)} = c_d\beta^{1/(d-1)}|A|^{d+1}.$$

If we compare this bound to (4.1), and choose β such that $\frac{\gamma\alpha}{2d} = c_d\beta^{1/(d-1)}$, then at least half of the popular points are covered by $d+1$ β-rich hyperplanes parallel to the facets of S_d.

Finally we can apply Lemma 2.3 with the pointset P of 'well-covered' popular points of $\mathbb{A}_d$ and with the sets of hyperplanes $L = \bigcup_{f \subset S_d} \mathscr{H}_f$. The number of points is at least

$\frac{\gamma\alpha}{2}|A|^d$. For a given f, $|\mathscr{H}_f| \leqslant \frac{|A|^d}{\beta|A|^{d-1}} = |A|/\beta$. The number of hyperplanes in L is at most $(d+1)|A|/\beta$. By Lemma 2.3, we have at least $\delta'|A|^{d+1}$ homothetic copies of S_d in $\mathbb{A}_d$. Let us project them onto x_1, the first coordinate axis. Every image is a $(k+1)$-term arithmetic progression, and the multiplicity of one image is at most $|A|^{d-1}$. Therefore there are at least $\delta'|A|^2$ $(k+1)$-term arithmetic progressions in A.

5. $G_n = K_n$

When the full sumset $A+A$ is small then it is easier to prove that A contains long arithmetic progressions. We can use the following Plünnecke-type inequality [8, 9, 12].

Theorem 5.1. *Let A and B be finite subsets of an abelian group such that $|A| = n$ and $|A+B| \leqslant \delta n$. Let $k \geqslant 1$ and $l \geqslant 1$. Then*

$$|kB - lB| \leqslant \delta^{k+l} n.$$

It follows from the inequality that, for any dimension d and d-dimensional integer vector $\vec{v} = (x_1, \ldots, x_d)$, $x_i \in \mathbb{Z}$, there is a $c > 0$ depending on d, δ and $\vec{v}$ such that the following holds: *If $|A+A| \leqslant \delta|A|$, then $\mathbb{A}_d$ can be covered by $c|A|$ hyperplanes with the same normal vector $\vec{v}$.* Using this, we can define our hyperplane-point arrangement, with the hyperplanes parallel to the facets of S_d containing at least one point of $\mathbb{A}_d$, and the pointset of the arrangement is $\mathbb{A}_d$. Then we do not have to deal with rich planes and popular points, and we can apply Lemma 2.3 directly.

References

[1] Balog, A. and Szemerédi, E. (1994) A statistical theorem of set addition. *Combinatorica* **14** 263–268.

[2] Elekes, G. (2002) Sums versus products in number theory, algebra and Erdős geometry. In *Paul Erdős and his Mathematics II*, Budapest, Vol. 11 of *Bolyai Society Mathematical Studies*, p. 277.

[3] Elekes, G. and Erdős, P. (1994) Similar configurations and pseudo grids. In *Intuitive Geometry*, Vol. 63 of *Coll. Math. Soc. János Bolyai*, North-Holland, Amsterdam, pp. 85–104.

[4] Frankl, P. and Rödl, V. (2002) Extremal problems on set systems. *Random Struct. Alg.* **20** 131–164.

[5] Freiman, G. A. (1973) *Foundations of Structural Theory of Set Addition*, Vol. 37 of *Translation of Mathematical Monographs*, AMS, Providence, RI, USA.

[6] Gowers, W. T. Hypergraph regularity and the multidimensional Szemerédi theorem. Manuscript.

[7] Nagle, B., Rödl, V. and Schacht, M. The counting lemma for regular k-uniform hypergraphs. *Random Struct. Alg.* **28** 113–179.

[8] Nathanson, M. B. (1996) *Additive Number Theory: Inverse Problems and the Geometry of Sumsets*, Vol. 165 of *Graduate Texts in Mathematics*, Springer.

[9] Plünnecke, H. (1969) *Eigenschaften und Abschätzungen von Wirkungsfunctionen*, Vol. 22, Berichte der Gesellschaft für Mathematik und Datenverarbeitung, Bonn.

[10] Rödl, V. and Skokan, J. (2004) Regularity lemma for k-uniform hypergraphs. *Random Struct. Alg.* **25** 1–42.

[11] Roth, K. F. (1953) On certain sets of integers. *J. London Math. Soc.* **28** 245–252.

[12] Ruzsa, I. Z. (1989) An application of graph theory to additive number theory. *Scientica Ser. A* **3** 97–109.

[13] Ruzsa, I. Z. and Szemerédi, E. (1978) Triple systems with no six points carrying three triangles. In Vol. 18 of *Colloquia Mathematica Societatis János Bolyai* (Proc. 5th Hungarian Colloq., Keszthely, Hungary, 1976), North-Holland, pp. 939–945.
[14] Solymosi, J. (2004) A note on a question of Erdős and Graham. *Combin. Probab. Comput.* **13** 263–267.
[15] Szemerédi, E. (1975) On sets of integers containing no k elements in arithmetic progression. *Acta Arithmetica* **27** 199–245.

Journal de Théorie des Nombres
de Bordeaux **17** (2005), 921–924

On sum-sets and product-sets of complex numbers

par JÓZSEF SOLYMOSI

RÉSUMÉ. On donne une preuve simple que pour tout ensemble fini de nombres complexes A, la taille de l'ensemble de sommes $A+A$ ou celle de l'ensemble de produits $A\cdot A$ est toujours grande.

ABSTRACT. We give a simple argument that for any finite set of complex numbers A, the size of the the sum-set, $A+A$, or the product-set, $A\cdot A$, is always large.

1. Introduction

Let A be a finite subset of complex numbers. The *sum-set* of A is $A+A=\{a+b : a,b\in A\}$, and the *product-set* is given by $A\cdot A=\{a\cdot b : a,b\in A\}$. Erdős conjectured that for any n-element set the sum-set or the product-set should be close to n^2. For integers, Erdős and Szemerédi [7] proved the lower bound $n^{1+\varepsilon}$.

$$\max(|A+A,|A\cdot A|)\geq |A|^{1+\varepsilon}.$$

Nathanson [9] proved the bound with $\varepsilon=1/31$, Ford [8] improved it to $\varepsilon=1/15$, and the best bound is obtained by Elekes [6] who showed $\varepsilon=1/4$ if A is a set of real numbers. Very recently Chang [3] proved $\varepsilon=1/54$ to finite sets of complex numbers. For further results and related problems we refer to [4, 5] and [1, 2].

In this note we prove Elekes' bound for complex numbers.

Theorem 1.1. *There is a positive absolute constant c, such that, for any finite sets of complex numbers A,B, and Q,*

$$c|A|^{3/2}|B|^{1/2}|Q|^{1/2}\leq |A+B|\cdot|A\cdot Q|,$$

whence $c|A|^{5/4}\leq\max\{|A+A|,|A\cdot A|\}$.

Manuscrit reçu le 26 aout 2003.
This research was supported by an NSERC grant.

2. Proof

For the proof we need some simple observations and definitions. For each $a \in A$ let us find "the closest" element, an $a' \in A$ so that $a' \neq a$ and for any $a'' \in A$ if $|a - a'| > |a - a''|$ then $a = a''$. If there are more then one closest elements, then let us select any of them. This way we have $|A|$ ordered pairs, let us call them *neighboring pairs*.

Definition. We say that a quadruple (a, a', b, q) is *good* if (a, a') is a neighboring pair, $b \in B$ and $q \in Q$, moreover

$$\left|\{u \in A + B : |a + b - u| \leq |a - a'|\}\right| \leq \frac{28|A + B|}{|A|}$$

and

$$\left|\{v \in A \cdot Q : |aq - v| \leq |aq - a'q|\}\right| \leq \frac{28|A \cdot Q|}{|A|}.$$

When a quadruple (a, a', b, q) is good, then it means that the neighborhoods of $a + b$ and aq are not very dense in $A + B$ and in $A \cdot Q$.

Lemma 2.1. *For any $b \in B$ and $q \in Q$ the number of good quadruples (a, a', b, q) is at least $|A|/2$.*

Proof. Let us consider the set of disks around the elements of A with radius $|a - a'|$ (i.e. for every $a \in A$ we take the largest disk with center a, which contains no other elements of A in it's interior). A simple geometric observation shows that no complex number is covered by more then 7 disks. Therefore

$$\sum_{a \in A} \left|\{u \in A + B : |a + b - u| \leq |a - a'|\}\right| \leq 7|A + B|$$

and

$$\sum_{a \in A} \left|\{v \in A \cdot Q : |aq - v| \leq |aq - a'q|\}\right| \leq 7|A \cdot Q|$$

providing that at least half of the neighboring pairs form good quadruples with b and q. Indeed, if we had more then a quarter of the neighboring pairs so that, say,

$$\left|\{v \in A \cdot Q : |aq - v| \leq |aq - a'q|\}\right| > \frac{28|A \cdot Q|}{|A|}$$

then it would imply

$$7|A \cdot Q| \geq \frac{|A|}{4}\left|\{v \in A \cdot Q : |aq - v| \leq |aq - a'q|\}\right| > 7|A \cdot Q|.$$

□

Proof of Theorem 1 To prove the theorem, we count the good quadruples (a, a', b, q) twice. For the sake of simplicity let us suppose that $0 \notin Q$. Such a quadruple is uniquely determined by the quadruple $(a+b, a'+b, aq, a'q)$. Now observe that there are $|A+B|$ possibilities for the first element, and given the value of $a+b$, the second element $a'+b$ must be one of the $28|A+B|/|A|$ nearest element of the sum-set $A+B$. We make the same argument for the third and fourth component to find that the number of such quadruples is at most

$$|A+B|\frac{28|A+B|}{|A|}|A\cdot Q|\frac{28|A\cdot Q|}{|A|}.$$

On the other hand, by Lemma 1 the number of such quadruples is at least

$$\frac{|A|}{2}|B||Q|$$

that proves the theorem.

A similar argument works for quaternions and for other hypercomplex numbers. In general, if T and Q are sets of similarity transformations and A is a set of points in space such that from any quadruple $(t(p_1), t(p_2), q(p_1), q(p_2))$ the elements $t \in T$, $q \in Q$, and $p_1 \neq p_2 \in A$ are uniquely determined, then

$$c|A|^{3/2}|T|^{1/2}|Q|^{1/2} \leq |T(A)|\cdot|Q(A)|,$$

where c depends on the dimension of the space only.

References

[1] J. Bourgain,S. Konjagin, *Estimates for the number of sums and products and for exponential sums over subgrups in finite fields of prime order*. C. R. Acad. Sci. Paris, (to appear).

[2] J. Bourgain, N. Katz, T. Tao, *A sum-product estimate in finite fields, and applications*. Geometric And Functional Analysis (to appear).

[3] M. Chang, *A sum-product estimate in algebraic division algebras over R*. Israel Journal of Mathematics (to appear).

[4] M. Chang, *Factorization in generalized arithmetic progressions and applications to the Erdős-Szemerédi sum-product problems*. Geometric And Functional Analysis (to appear).

[5] M. Chang, *Erdős-Szemerédi sum-product problem*. Annals of Math. **157** (2003), 939–957.

[6] Gy. Elekes, *On the number of sums and products*. Acta Arithmetica **81** (1997), 365–367.

[7] P. Erdős, E. Szemerédi, *On sums and products of integers*. In: Studies in Pure Mathematics; To the memory of Paul Turán. P.Erdős, L.Alpár, and G.Halász, editors. Akadémiai Kiadó – Birkhauser Verlag, Budapest – Basel-Boston, Mass. 1983, 213–218.

[8] K. Ford, *Sums and products from a finite set of real numbers*. Ramanujan Journal, **2** (1998), (1-2), 59–66.

[9] M. B. Nathanson, *On sums and products of integers*. Proc. Am. Math. Soc. **125** (1997), (1-2), 9–16.

József Solymosi
Department of Mathematics,
University of British Columbia
1984 Mathematics Road, Vancouver,
Colombie-Britannique, Canada V6T 1Z2
E-mail : solymosi@math.ubc.ca

Bull. London Math. Soc. 37 (2005) 491–494
doi:10.1112/S0024609305004261

ON THE NUMBER OF SUMS AND PRODUCTS

JÓZSEF SOLYMOSI

Abstract

A new lower bound on $\max\{|A+A|, |A\cdot A|\}$ is given, where A is a finite set of complex numbers.

1. *Introduction*

Let A be a finite subset of complex numbers. The *sum-set* of A is $A+A=\{a+b : a,b\in A\}$, and the *product-set* is $A\cdot A=\{a\cdot b : a,b\in A\}$. Erdős and Szemerédi [**7**] proved the inequality

$$\max(|A+A|, |A\cdot A|) \geqslant c|A|^{1+\varepsilon}$$

for a small but positive ε, where A is a subset of integers. They conjectured that

$$\max(|A+A|, |A\cdot A|) \geqslant c|A|^{2-\delta}$$

for any positive δ. (In this paper, c stands for the general constant. Some authors use the $n \ll m$ or $n \gg m$ notation instead of our $n \leqslant cm$ or $n \geqslant cm$.)

After improvements given in [**9**], [**8**], and [**3**], the best bound so far has been obtained by Elekes [**4**], who showed that $\varepsilon \geqslant 1/4$ if A is a set of real numbers. His result was extended to complex numbers in [**13**] and [**11**]. For further results and related problems, we refer the reader to [**1**] and [**5**].

In this paper, we prove the following theorem.

Theorem 1. *There is a positive absolute constant c such that, for every n-element set A,*

$$\frac{cn^{14}}{\log^3 n} \leqslant |A+A|^8 \cdot |A\cdot A|^3,$$

whence $cn^{14/11}/\log^{3/11} n \leqslant \max\{|A+A|, |A\cdot A|\}$.

Nathanson and Tenenbaum [**10**] proved that the product set should be large, namely $|A|^{2-\varepsilon}$, if the sumset is at most $3|A|-4$. Chang [**2**], and independently Elekes and Ruzsa [**6**], proved a similar bound if the sumset is at most $c|A|$. As a consequence of Theorem 1, we obtain the following corollary.

Corollary 1. *If $|A|=n$ and $|A+A| \leqslant Cn$, then $|A\cdot A| \geqslant cn^2/\log n$.*

Received 1 August 2003; revised 8 June 2004.

2000 *Mathematics Subject Classification* 11B75 (primary), 52C10 (secondary).

This research was supported by NSERC and OTKA grants.

2. *Proof*

Our proof is based on the following estimates of the number of incidences between lines and points.

THEOREM 2 (Szemerédi and Trotter [**12**]). *The maximum number of incidences between n points and m straight lines of the real plane is $O(n^{2/3}m^{2/3}+n+m)$.*

COROLLARY 2 (Szemerédi and Trotter [**12**]). *Given a set of n points on the real plane, the number of k-rich lines (that is, lines incident to at least k points) is $O(n^2/k^3+n/k)$.*

In the proof of Theorem 1 we use Theorem 2 and Corollary 2 on Cartesian products only; similar statements are easy to prove for complex lines in the complex plane. (The general case has recently been solved by Tóth [**13**].) The following lemma has been proved but not published by the author.

LEMMA 1. *Given two sets of complex numbers S_1 and S_2 with sizes $|S_1| = n_1$ and $|S_2| = n_2$, let $S = S_1 \times S_2$ be the Cartesian product. The maximum number of incidences between S and m complex lines of the complex plane is $O((n_1n_2)^{2/3}m^{2/3}+n_1n_2+m)$.*

Proof of Theorem 1. If $|A\cdot A| = t$, then the number of pairs $(a_i,a_j),(a_u,a_v)$ such that $a_i\cdot a_j = a_u\cdot a_v$ (where $a_i,a_j,a_u,a_v\in A$) is at least cn^4/t. Then the number of pairs $(a_i,a_v),(a_u,a_j)\in A\times A$, where $a_i/a_v = a_u/a_j$, is at least cn^4/t as well. Let us partition the elements of $A\times A$ into classes (lines) $L_1,L_2,\dots,L_k$ using the relation $(a_i,a_j)\sim(a_u,a_v)$ if and only if $a_i/a_j = a_u/a_v$. Each class is a collection of collinear points, and the line through them contains the origin $(0,0)$. If l_i denotes the size of L_i, then

$$\sum_{i=1}^{k}\binom{l_i}{2}\geqslant\frac{cn^4}{t}.$$

We partition these lines into sets $C_1,C_2,\dots,C_s$ ($s\leqslant\log n^2$) with respect to their 'squared' sizes. Then $L_i\in C_j \iff 2^{2(j-1)} < \binom{l_i}{2}\leqslant 2^{2j}$. There are at most $\log n^2$ sets, so there is at least one set, C_j, which covers many elements. Let X_j be a set of all pairs $((a_\nu,a_\mu),(a_\varrho,a_\rho))$ such that there exists L_i in C_j with (a_ν,a_μ) and (a_ϱ,a_ρ) both in L_i. Then at least one of the sets X_j is large. Also,

$$|X_j| = |\{(a_\nu,a_\mu),(a_\varrho,a_\rho) : (a_\nu,a_\mu)\in L_i,(a_\varrho,a_\rho)\in L_i, L_i\in C_j\}|\geqslant\frac{cn^4}{t\log n},$$

and therefore

$$2^{2j}|C_j|\geqslant\frac{cn^4}{t\log n}. \tag{2.1}$$

This is the key element of the proof: every point of $A\times A$ is incident to at least $|C_j|$ lines, each of them incident to at least 2^{j-1} points of $(A+A)\times(A+A)$. Indeed, the translated lines $(a_u,a_v)+L$ with L in C_j are incident to (a_u,a_v), and the points of the lines are points from $(A+A)\times(A+A)$ (see Figure 1). We denote the set of translated lines by $\mathcal{L}$, as follows:

$$\mathcal{L} = \{(a_u,a_v)+L : L\in C_j,(a_u,a_v)\in A\times A\}.$$

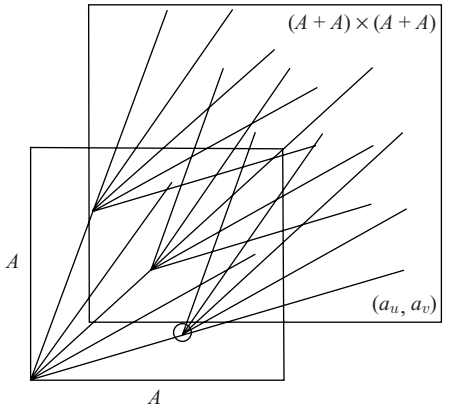


FIGURE 1. *Translates of the lines of* C_j.

Because of Corollary 2, the number of 2^{j-1}-rich lines (that is, lines incident to at least 2^{j-1} points) on $(A+A)\times(A+A)$ is

$$O\left(\frac{|A+A|^4}{(2^{j-1})^3}+\frac{|A+A|}{(2^{j-1})}\right).$$

The first term is always larger than the second because $|A+A|>|A|$ and $2^{j-1}\leqslant|A|$. Therefore,

$$|\mathcal{L}|\leqslant\frac{c|A+A|^4}{(2^{j-1})^3}.$$

Applying the bound from Theorem 2 to the number of incidences I between $\mathcal{L}$ and the n^2 points of $A\times A$, we have

$$I=O\big(|\mathcal{L}|^{2/3}(n^2)^{2/3}+|\mathcal{L}|+n^2\big).$$

Therefore,

$$n^2|C_j|\leqslant c|\mathcal{L}|^{2/3}n^{4/3}, \tag{2.2}$$

or

$$n^2|C_j|\leqslant c|\mathcal{L}|, \tag{2.3}$$

or

$$n^2|C_j|\leqslant cn^2. \tag{2.4}$$

The right-hand side of (2.2) is always at least cn^2, and therefore (2.2) includes case (2.4). The next step is to see that (2.2) covers case (2.3) as well. Let us suppose that, on the contrary,

$$|\mathcal{L}|^{2/3}n^{4/3}<|\mathcal{L}|.$$

Then

$$n^{4/3}<|\mathcal{L}|^{1/3}\to n^4<|\mathcal{L}|,$$

but this is not possible, since $\mathcal{L}$ consists of n^2 translates of less than n^2 lines.

Now we are ready for the final step of the proof. It follows from (2.1), that

$$2^{2j-2} \geqslant \frac{cn^4}{t \log n |C_j|}. \tag{2.5}$$

Putting (2.2) and (2.5) together, we have

$$\begin{aligned} n^2|C_j| &\leqslant c|\mathcal{L}|^{2/3} n^{4/3} \leqslant c\left(\frac{|A+A|^4}{(2^{j-1})^3}\right)^{2/3} n^{4/3} = c\frac{|A+A|^{8/3}}{2^{2j-2}} n^{4/3} \\ &\leqslant c\frac{|A+A|^{8/3}}{(n^4/t\log n|C_j|)} n^{4/3}, \end{aligned}$$

which gives

$$\frac{cn^{14}}{\log^3 n} \leqslant |A+A|^8 \cdot t^3,$$

as stated. □

Acknowledgement. The author wishes to thank the referee for helping to improve the style and clarity of the paper.

References

1. J. Bourgain, K. Katz and T. Tao, 'A sum-product estimate in finite fields, and applications', *Geom. Funct. Anal.* 14 (2004) 27–57.
2. M. Chang, 'Factorization in generalized arithmetic progressions and applications to the Erdős–Szemerédi sum-product problems', *Geom. Funct. Anal.* 13 (2003) 720–736.
3. M.-C. Chang, 'A sum-product estimate in algebraic division algebras', *Israel J. Math.*, to appear.
4. Gy. Elekes, 'On the number of sums and products', *Acta Arith.*, 81 (1997) 365–367.
5. Gy. Elekes, 'Sums versus products in number theory, algebra and Erdős geometry', *Paul Erdős and his mathematics, II* (*Budapest*, 1999), Bolyai Soc. Math. Stud. 11 (ed. G. Halász, János Bolyai Math. Soc., Budapest, 2002) 241–290.
6. Gy. Elekes and I. Z. Ruzsa, 'Few sums, many products', *Studia Sci. Math. Hungar.* 40 (2003) 301–308.
7. P. Erdős and E. Szemerédi, 'On sums and products of integers', *Studies in Pure Mathematics* (Birkhäuser, Basel, 1983) 213–218.
8. K. Ford, 'Sums and products from a finite set of real numbers', *Ramanujan J.* 2 (1998) 59–66.
9. M. B. Nathanson, 'On sums and products of integers', *Proc. Amer. Math. Soc.* 125 (1997) 9–16.
10. M. B. Nathanson and G. Tenenbaum, 'Inverse theorems and the number of sums and products', *Astérisque* 258 (1999) 195–204.
11. J. Solymosi, 'On sums and products of complex numbers', *J. Théor. Nombres Bordeaux*, to appear.
12. E. Szemerédi and W. T. Trotter, Jr., 'Extremal problems in discrete geometry', *Combinatorica* 3 (1983) 381–392.
13. C. D. Tóth, 'The Szemerédi–Trotter theorem in the complex plane', *Math ArXiV*, eprint:math/0305283, 2003.

József Solymosi
Department of Mathematics
The University of British Columbia
1984 Mathematics Road
Vancouver, B.C.
Canada V6T 1Z2

solymosi@math.ubc.ca

dc_52_10
Bull. London Math. Soc. 41 (2009) 817–822
doi:10.1112/blms/bdp054

Sum–product estimates for well-conditioned matrices

J. Solymosi and V. Vu

Dedicated to the memory of György Elekes

Abstract

We show that if $\mathcal{A}$ is a finite set of $d \times d$ well-conditioned matrices with complex entries, then the following sum–product estimate holds $|\mathcal{A}+\mathcal{A}| \times |\mathcal{A}\cdot\mathcal{A}| = \Omega(|\mathcal{A}|^{5/2})$.

1. *Introduction*

Let $\mathcal{A}$ be a finite subset of a ring Z. The sum–product phenomenon, first investigated by Erdős and Szemerédi [4], suggests that either $\mathcal{A}\cdot\mathcal{A}$ or $\mathcal{A}+\mathcal{A}$ is much larger than $\mathcal{A}$. This was first proved for $\mathbb{Z}$, the ring of integers, in [4]. Recently, many researchers have studied (with considerable success) other rings. Several of these results have important applications in various fields of mathematics. The interested readers are referred to Bourgain's survey [1].

In this paper we consider Z being the ring of $d \times d$ matrices with complex entries. (We are going to use the notation 'matrix of size d' for $d \times d$ matrices.) It is well known that one cannot generalize the sum–product phenomenon, at least in the straightforward manner, in this case. The archetypal counterexample is the following:

Example 1.1. Let I denote the identity matrix and let E_{ij} be the matrix with only one non-zero entry at position ij and this entry is one. Let $M_a := I + aE_{1d}$ and let $\mathcal{A} = \{M_1, \dots, M_n\}$. It is easy to check that $|\mathcal{A}+\mathcal{A}| = |\mathcal{A}\cdot\mathcal{A}| = 2n-1$.

This example suggests that one needs to make some additional assumptions in order to obtain a non-trivial sum–product estimate. Chang [2] proved the following

Theorem 1.2. *There is a function $f = f(n)$ tending to infinity with n such that the following holds. Let $\mathcal{A}$ be a finite set of matrices of size d over the reals such that for any $M \neq M' \in \mathcal{A}$, we have $\det(M-M') \neq 0$. Then we have*

$$|\mathcal{A}+\mathcal{A}| + |\mathcal{A}\cdot\mathcal{A}| \geqslant f(|\mathcal{A}|)|\mathcal{A}|.$$

The function f in Chang's proof tends to infinity slowly. In most applications, it is desirable to have a bound of the form $|\mathcal{A}|^{1+c}$ for some positive constant c. In this paper, we show that this is indeed the case (and in fact c can be set to be $\frac{1}{4}$) if we assume that the matrices are far from being singular. Furthermore, this result provides a new insight into the above counterexample (see the discussion following Theorem 2.2).

Received 12 February 2008; revised 9 April 2009; published online 19 July 2009.

2000 *Mathematics Subject Classification* 11B75 (primary), 15A45, 11C20 (secondary).

The research was conducted while both researchers were members of the Institute for Advanced Study. Funding provided by The Charles Simonyi Endowment. The first author was supported by NSERC and OTKA grants and by Sloan Research Fellowship. The second author was supported by an NSF Career Grant.

NOTATION. We use asymptotic notation under the assumption that $|\mathcal{A}| = n$ tends to infinity. Notation such as $f(n) = \Omega_\xi(m)$ means that there is a constant $c > 0$, which depends on ξ only, such that $f(n) \geqslant cm$ for every large enough n. Throughout the paper letter ξ might be a number like d or a vector like κ, d or α, r. The notation $f(n) = \mathrm{O}_\xi(m)$ means that there is a constant c, which depends on ξ only, such that $f(n) \leqslant cm$ for every large enough n. In both cases m is a function of n or it is the constant one function, $m = 1$, in which case we write $\Omega_\xi(1)$ or $\mathrm{O}_\xi(1)$. Throughout the paper symbol $\mathbb{C}$ denotes the field of complex numbers.

2. *New results*

The classical way to measure how close a matrix is to being singular is to consider its *condition* number.

For a matrix M of size d, let $\sigma_{\max}(M)$ and $\sigma_{\min}(M)$ be the largest and smallest singular values of M. The quantity $\kappa(M) = \sigma_{\max}(M)\sigma_{\min}(M)^{-1}$ is the *condition* number of M. (If M is singular, then $\sigma_{\min}(M) = 0$ and $\kappa(M) = \infty$.)

Our main result shows that if the matrices in $\mathcal{A}$ are well conditioned (that is, their condition numbers are small, or equivalently they are far from being singular), then $|\mathcal{A} + \mathcal{A}| + |\mathcal{A} \cdot \mathcal{A}|$ is large.

DEFINITION 2.1. Let κ be a positive number at least one. A set $\mathcal{A}$ of matrices is called κ-well conditioned if the following conditions hold.

(i) For any $M \in \mathcal{A}$, we have $\kappa(M) \leqslant \kappa$.
(ii) For any $M, M' \in \mathcal{A}$, we have $\det(M - M') \neq 0$, unless $M = M'$.

THEOREM 2.2. *Let $\mathcal{A}$ be a finite κ-well-conditioned set of size d matrices with complex entries. Then we have*

$$|\mathcal{A} + \mathcal{A}| \times |\mathcal{A} \cdot \mathcal{A}| \geqslant \Omega_{\kappa,d}(|\mathcal{A}|^{5/2}).$$

Consequently, we have

$$|\mathcal{A} + \mathcal{A}| + |\mathcal{A} \cdot \mathcal{A}| \geqslant \Omega_{\kappa,d}(|\mathcal{A}|^{5/4}).$$

Theorem 2.2 is a generalization of the first author's sum–product bound on complex numbers [7]. Some elements in the proof of Theorem 2.2 were inspired by techniques applied in [7]. The idea of using geometry for sum–product problems was introduced by Elekes [3].

REMARK 2.3. By following the proof closely, one can set the hidden constant in Ω as $(\frac{c}{\kappa})^{d^2}$, where c is an absolute constant ($\frac{1}{100}$, say, would be sufficient).

REMARK 2.4. We reconsider the set in the counterexample. It is easy to show that both $\sigma_{\max}(M_a)$ and $\sigma_{\min}(M_a)^{-1}$ are $\Omega_d(a)$. Thus $\kappa(M_a) = \Omega_d(a^2)$, which, for a typical a, is $\Omega_d(|\mathcal{A}|^2)$. Hence, the matrices in the counterexample have very large condition numbers.

REMARK 2.5. Note that if the entries of a matrix M of size d are random integers from $\{-n, \dots, n\}$, then, with probability tending to one as n tends to infinity, $\kappa(M) = \mathrm{O}_d(1)$. (In order to see this, note that by Hadamard's bound, $\sigma_{\max}(M) \leqslant dn$ with probability one. Moreover, it is easy to show that with high probability $|\det M| = \Omega_d(n^d)$, which implies that $\sigma_{\min}(M) = \Omega_d(n)$.)

The proof of Theorem 2.2 is presented in Sections 3–6.

3. *Neighborhoods*

Consider a matrix M of size d. We can view M as a vector in $\mathbb{C}^{d^2}$ by writing its entries (from left to right, row to row) as the co-ordinates. From now on we consider $\mathcal{A}$ as a subset of $\mathbb{C}^{d^2}$. The matrix operations act as follows:

(i) *addition*: this will be viewed as vector addition;

(ii) *multiplication*: this is a bit more tricky. Take a matrix M of size d and a d^2-vector M'. To obtain the vector $M'M$, we first rewrite M' as a matrix, then do the matrix multiplication $M'M$, and finally rewrite the result as a vector. This multiplying by M is a linear operator on $\mathbb{C}^{d^2}$.

Next, we need a series of definitions. Note that here we are considering M as a vector in $\mathbb{C}^{d^2}$. The norm $\|M\|$ indicates the length of this vector in $\mathbb{C}^{d^2}$. Then we have the following.

(i) Radius of M, that is, $r(M) := \min_{M' \in \mathcal{A} \setminus \{M\}} \|M - M'\|$.

(ii) Nearest neighbor of M, that is, $n(M)$ is an M' such that $\|M - M'\| = r(M)$ (if there is more than one M' then choose one arbitrarily).

(iii) Ball of M, that is, $B(M)$ is the ball in $\mathbb{C}^{d^2}$ around M with radius $r(M)$.

The following lemma will be used frequently in the proof. Let x, y, z be three different points in $\mathbb{C}^r$. The angle xyz is the angle between the rays yx and yz. We understand that this angle is at most π. In $\mathbb{C}^r$ there are various ways of defining the angle between two vectors x and y. (See [6] for a survey of some possible choices.) We are using the

$$\angle(x, y) = \arccos \frac{\mathrm{Re}(y^* x)}{\|x\|\|y\|}$$

notation, where $\mathrm{Re}(y^* x)$ is the real part of the Hermitian product, $(y^* x) = \sum_{i=1}^{r} \bar{y}_i x_i$. It is important to us that with this definition the law of cosines remains valid, and we have

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| \cos(\angle(x, y)). \tag{3.1}$$

LEMMA 3.1. *For any positive integer r and any constant $0 < \alpha \leqslant \pi$, there is a constant $C(\alpha, r)$ such that the following holds. There are at most $C(\alpha, r)$ points on the unit sphere in $\mathbb{C}^r$ such that for any two points z, z', the angle zoz' is at least α. (Here o denotes the origin.)*

This lemma is equivalent to the statement that a unit sphere in $\mathbb{C}^r$ has at most $C(\delta, r)$ points such that any two has distance at least δ. It can be proved using a simple volume argument. (See [5] for a more advanced approach.) The optimal estimate for $C(\alpha, r)$ is unknown for most pairs (α, r), but this value is not important in our argument.

LEMMA 3.2. *For any positive integer r there is a positive constant $C_1(r)$ such that the following holds. Let $\mathcal{A}$ be a set of points in $\mathbb{C}^r$. Then for $z \in \mathbb{C}^r$ there are at most $C_1(r)$ elements M of $\mathcal{A}$ such that $z \in B(M)$.*

Proof. Let $M_1, \ldots, M_k$ be elements of $\mathcal{A}$ such that $z \in B(M_i)$ for all i. By the definition of $B(M)$ the distance between two distinct elements, M_i and M_j, is at least as large as their distances from z. Then, by (3.1), the angle $M_i z M_j$ is at least $\pi/3$ for any $i \neq j$. The claim follows from Lemma 3.1. □

4. *K-normal pairs*

Let K be a large constant to be determined. We call an ordered pair (M, M') *product K-normal* if the ellipsoid $B(M)M'$ contains at most $K(|\mathcal{A}\cdot\mathcal{A}|/|\mathcal{A}|)$ points from $\mathcal{A}\cdot\mathcal{A}$. (Recall that multiplying by M' is a linear operator on $\mathbb{C}^{d^2}$, and thus it maps a ball into an ellipsoid.)

LEMMA 4.1. *There is a constant $C_2 = C_2(d)$ such that the following holds. For any fixed M' and $K \geqslant C_2$, the number of M such that the pair (M, M') is product K-normal is at least $(1 - C_2/K)|\mathcal{A}|$.*

Proof. Let $M_1, \dots, M_m$ be the elements of $\mathcal{A}$, where (M_i, M) is not product K-normal. By definition, we have

$$\sum_{i=1}^{m} |B(M_i)M \cap \mathcal{A}\cdot\mathcal{A}| \geqslant Km\frac{|\mathcal{A}\cdot\mathcal{A}|}{|\mathcal{A}|}.$$

Set $\varepsilon := m/|\mathcal{A}|$. By the pigeon hole principle, there is a point z in $\mathcal{A}\cdot\mathcal{A}$ belonging to at least $K\varepsilon$ ellipsoids $B(M_i)M$. By applying the map M^{-1}, it follows that zM^{-1} belongs to at least $K\varepsilon$ balls $B(M_i)$. By Lemma 3.2, $K\varepsilon = \mathrm{O}(d^2) = \mathrm{O}(d)$. Thus, $\varepsilon = \mathrm{O}(d)/K$, proving the claim. □

By the same argument, we can prove the sum version of this lemma. An ordered pair (M, M') is *sum K-normal* if the ball $B(M) + M'$ contains at most $K(|\mathcal{A}+\mathcal{A}|/|\mathcal{A}|)$ points from $\mathcal{A}+\mathcal{A}$.

LEMMA 4.2. *For any fixed M', the number of M such that the pair (M, M') is sum K-normal is at least $(1 - C_2/K)|\mathcal{A}|$.*

5. *Cones*

For a ball B in $\mathbb{C}^r$ and a point $x \notin B$, define the cone $\mathrm{Cone}(x, B)$ as

$$\mathrm{Cone}(x, B) := \{tx + (1-t)B | 0 \leqslant t \leqslant 1\}.$$

Now let α be a positive constant at most π. For two different points x and y, we define the cone $\mathrm{Cone}_\alpha(x, y)$ as $\mathrm{Cone}(x, B_\alpha(y))$, where $B_\alpha(y)$ is the unique ball around y such that the angle of $\mathrm{Cone}(x, B_\alpha(y))$ is exactly α. (The angle of $\mathrm{Cone}(x, B_\alpha(y))$ is given by $\max_{s,t\in B_\alpha(y)} \angle sxt$.)

LEMMA 5.1. *For any positive integer r and any constant $0 < \alpha \leqslant \pi$, there is a constant $C(\alpha, r)$ such that the following holds. Let $\mathcal{A}$ be a finite set of points in $\mathbb{C}^r$ and let L be any positive integer. Then for any point $x \in \mathbb{C}^r$, there are at most $C(\alpha, r)L$ points y in $\mathcal{A}$ such that the cone $\mathrm{Cone}_\alpha(x, y)$ contains at most L points from $\mathcal{A}$.*

Proof. *Case* 1: We first prove the case $L = 1$. In this case, if $y \in \mathcal{A}$ and $\mathrm{Cone}_\alpha(x, y)$ contains at most one point from $\mathcal{A}$, then it contains exactly one point which is y. For any two points $y_1, y_2 \in \mathcal{A}$ such that both $\mathrm{Cone}_\alpha(x, y_1)$ and $\mathrm{Cone}_\alpha(x, y_2)$ contain exactly one point from $\mathcal{A}$, the angle y_1xy_2 is at least α, by the definition of the cones. Thus, the claim follows from Lemma 3.1.

Case 2 : We reduce the case of general L to the case $L = 1$ by a random sparsifying argument. Let $\mathcal{Y} = \{y_1, \dots, y_m\}$ be a set of points in $\mathcal{A}$ such that $\mathrm{Cone}_\alpha(x, y_i)$ contains at

most L points from $\mathcal{A}$ for all $1 \leqslant i \leqslant m$. We create a random subset $\mathcal{A}'$ of $\mathcal{A}$ by picking each point with probability p (for some $0 < p \leqslant 1$ to be determined), randomly and independently. We say that y_i *survives* if it is chosen and no other points in $\mathcal{A} \cap \mathrm{Cone}_\alpha(x, y_i)$ are chosen. For each $y_i \in \mathcal{Y}$, the probability that it survives is at least $p(1-p)^{L-1}$. By linearity of expectations, the expected number of points that survive is at least $mp(1-p)^L$. Thus, there are sets $\mathcal{Y}' \subset \mathcal{A}' \subset \mathcal{A}$, where $|\mathcal{Y}'| \geqslant mp(1-p)^L$ with the property that each point $y_i \in \mathcal{Y}'$ is the only point in $\mathcal{A}'$ that appears in $\mathrm{Cone}(x, y_i) \cap \mathcal{A}'$. By the special case $L = 1$, we conclude that $mp(1-p)^{L-1} \leqslant |\mathcal{Y}'| = \mathrm{O}_{\alpha,r}(1)$. The claim of the lemma follows by setting $p = 1/L$. □

6. *Proof of the main theorem*

Consider a point M and its nearest neighbor $n(M)$. Let M_1 be another point, viewed as a matrix. We consider the multiplication with M_1. This maps the ball $B(M)$ to the ellipsoid $B(M)M_1$ and $n(M)$ to the point $n(M)M_1$.

Since the condition number $\kappa(M_1)$ is not too large, it follows that $B(M)M_1$ is not degenerate. In other words, the ratio between the maximum and minimum distance from MM_1 to a point on the boundary of $B(M)M_1$ is bounded from above by $\mathrm{O}_\kappa(1)$.

Let $b(M, M_1)$ be the largest ball contained in $B(M)M_1$ and $\mathrm{Cone}(M, M_1)$ be the cone with its tip at $n(M)M_1$ defined by

$$\mathrm{Cone}(M, M_1) := \{tn(M)M_1 + (1-t)b(M, M_1) | 0 \leqslant t \leqslant 1\}.$$

The assumption that M_1 is well conditioned implies that the angle of this cone is bounded from below by a positive constant α depending only on κ and d. Thus, we can apply Lemma 5.1 to this system of cones.

Let T be the number of ordered triples (M_0, M_1, M_2) such that (M_0, M_1) is product K-normal and (M_0, M_2) is sum K-normal.

We choose K sufficiently large so that the constant $(1 - C_2/K)$ in Lemmas 4.1 and 4.2 is at least $\frac{9}{10}$. It follows that for any fixed M_1 and M_2, there are at least $\frac{4}{5}|\mathcal{A}|$ matrices M_0 such that (M_0, M_1) is product K-normal and (M_0, M_2) is sum K-normal. This implies that

$$T \geqslant \frac{4}{5}|\mathcal{A}|^3. \tag{6.1}$$

Now we bound T from above. First we embed the triple (M_0, M_1, M_2) into the quadruple $(M_0, n(M_0), M_1, M_2)$. Next, we bound the number of $(M_0, n(M_0), M_1, M_2)$ from above.

The κ-well-conditioned assumption of Theorem 2.2 guarantees that the quadruple $(M_0, n(M_0), M_1, M_2)$ is uniquely determined by the quadruple

$$(M_0M_1, n(M_0)M_1, M_0 + M_2, n(M_0) + M_2).$$

In order to see this, set $A = M_0M_1, B = n(M_0)M_1, C = M_0 + M_2$ and $D = n(M_0) + M_2$. Then $(M_0 - n(M_0))M_1 = A - B$ and $M_0 - n(M_0) = C - D$. Since $M - M'$ is invertible for any $M \neq M' \in \mathcal{A}$, we have $M_1 = (C - D)^{-1}(A - B)$. (This is the only place where we use this condition.) Since M_1 is also invertible (as it has a bounded condition number), it follows that $M_0 = AM_1^{-1}$, $n(M_0) = BM_1^{-1}$ and $M_2 = C - M_0$.

It suffices to bound the number of $(M_0M_1, n(M_0)M_1, M_0 + M_2, n(M_0) + M_2)$.

We first choose $n(M_0)M_1$ from $\mathcal{A} \cdot \mathcal{A}$. There are, of course, $|\mathcal{A} \cdot \mathcal{A}|$ choices. After fixing this point, by Lemma 5.1 and the definition of product K-normality, we have $\mathrm{O}_{\kappa,d}(K(|\mathcal{A} \cdot \mathcal{A}|/|\mathcal{A}|))$ choices for M_0M_1. Similarly, we have $|\mathcal{A} + \mathcal{A}|$ choices for $n(M_0) + M_2$ and for each such choice, we have $O_{\kappa,d}(K(|\mathcal{A} + \mathcal{A}|/|\mathcal{A}|))$ choices for $M_0 + M_2$. It follows that

$$T \leqslant |\mathcal{A} \cdot \mathcal{A}| \cdot \mathrm{O}_{\kappa,d}\left(K\frac{|\mathcal{A} \cdot \mathcal{A}|}{|\mathcal{A}|}\right) \cdot |\mathcal{A} + \mathcal{A}| \cdot \mathrm{O}_{\kappa,d}\left(K\frac{|\mathcal{A} + \mathcal{A}|}{|\mathcal{A}|}\right). \tag{6.2}$$

Recall that K is also a constant depending only on κ and d. Putting (6.1) and (6.2) together, we obtain

$$\frac{4}{5}|\mathcal{A}|^3 \leqslant \mathrm{O}_{\kappa,d}\left(\frac{|\mathcal{A}\cdot\mathcal{A}||\mathcal{A}+\mathcal{A}|}{|\mathcal{A}|^2}\right),$$

concluding the proof. □

Acknowledgements. The authors thank an anonymous referee for useful comments on a previous draft.

References

1. J. Bourgain, 'More on the sum–product phenomenon in prime fields and its applications', *Int. J. Number Theory* 1 (2005) 1–32.
2. M.-C. Chang, 'Additive and multiplicative structure in matrix spaces', *Comb. Probab. Comput.* 16 (2007) 219–238.
3. Gy. Elekes, 'On the number of sums and products', *Acta Arith.* 81 (1997) 365–367.
4. P. Erdős and E. Szemerédi, 'On sums and products of integers', *Studies in pure mathematics* (Birkhauser, Basel, 1983) 213–218.
5. O. Henkel, 'Sphere-packing bounds in the Grassmann and Stiefel manifolds', *IEEE Trans. Inf. Theory* 51 (2005) 3445–3456.
6. K. Scharnhorst, 'Angles in complex vector spaces', *Acta Appl. Math.* 69 (2001) 95–103.
7. J. Solymosi, 'On sum-sets and product-sets of complex numbers', *J. Théor. Nombres Bordeaux* 17 (2005) 921–924.

J. Solymosi
Department of Mathematics
University of British Columbia
1984 Mathematics Road
Vancouver, BC
Canada V6T 1Z2

solymosi@math.ubc.ca

V. Vu
Department of Mathematics
Rutgers University
110 Frelinghuysen Road
Piscataway, NJ 08554
USA

vanvu@math.rutgers.edu

Available online at www.sciencedirect.com
ScienceDirect
Advances in Mathematics 222 (2009) 402–408
ADVANCES IN Mathematics
www.elsevier.com/locate/aim

Bounding multiplicative energy by the sumset ☆

József Solymosi

University of British Columbia, Mathematics, 1984 Mathematics Road, Vancouver, British Columbia, V6T 1Z2, Canada

Received 23 June 2008; accepted 15 April 2009

Available online 25 April 2009

Communicated by Gil Kalai

Abstract

We prove that the sumset or the productset of any finite set of real numbers, A, is at least $|A|^{4/3-\varepsilon}$, improving earlier bounds. Our main tool is a new upper bound on the multiplicative energy, $E(A, A)$.

Keywords: Sum-product estimates

1. Introduction

The *sumset* of a finite set of an additive group, A, is defined by

$$A + A = \{a + b\colon\ a, b \in A\}.$$

The *productset* and *ratioset* are defined in a similar way,

$$AA = \{ab\colon\ a, b \in A\},$$

and

$$A/A = \{a/b\colon\ a, b \in A\}.$$

A famous conjecture of Erdős and Szemerédi [5] asserts that for any finite set of integers, M,

☆ Research was supported by OTKA and NSERC grants and by a Sloan Fellowship.
E-mail address: solymosi@math.ubc.ca.

0001-8708/$ – see front matter
doi:10.1016/j.aim.2009.04.006

$$\max\{|M+M|, |MM|\} \geqslant |M|^{2-\varepsilon},$$

where $\varepsilon \to 0$ when $|M| \to \infty$. They proved that

$$\max\{|M+M|, |MM|\} \geqslant |M|^{1+\delta},$$

for some $\delta > 0$. In a series of papers, lower bounds on δ were find. $\delta \geqslant 1/31$ [10], $\delta \geqslant 1/15$ [6], $\delta \geqslant 1/4$ [3], and $\delta \geqslant 3/11$ [12]. The last two bonds were proved for finite sets of real numbers.

2. Results

Our main result is the following.

Theorem 2.1. *Let A be a finite set of positive real numbers. Then*

$$|AA||A+A|^2 \geqslant \frac{|A|^4}{4\lceil \log|A| \rceil}$$

holds.

The inequality is sharp—up to the power of the log term in the denominator—when A is the set of the first n natural numbers. Theorem 2.1 implies an improved bound on the sum-product problem.

Corollary 2.2. *Let A be a finite set of positive real numbers. Then*

$$\max\{|A+A|, |AA|\} \geqslant \frac{|A|^{4/3}}{2\lceil \log|A| \rceil^{1/3}}$$

holds

2.1. Proof of Theorem 2.1

To illustrate how the proof goes, we are making two unjustified and usually false assumptions, which are simplifying the proof. Readers, not interested in this "handwaving", will find the rigorous argument about 20 lines below.

Suppose that AA and A/A have the same size, $|AA| \approx |A/A|$, and any element of A/A has about the same number of representations as any other. This means that for any reals $s, t \in A/A$ the two numbers s and t have the same multiplicity, $|\{(a,b) \mid a,b \in A,\ a/b = s\}| \approx |\{(b,c) \mid b,c \in A,\ b/c = t\}|$. A geometric interpretation of the cardinality of A/A is that the Cartesian product $A \times A$ is covered by $|A/A|$ concurrent lines going through the origin. Label the rays from the origin covering the points of the Cartesian product anticlockwise by $r_1, r_2, \ldots, r_m$, where $m = |A/A|$.

Our assumptions imply that each ray is incident to $|A|^2/|AA|$ points of $A \times A$. Consider the elements of $A \times A$ as two-dimensional vectors. The sumset $(A \times A) + (A \times A)$ is the same set as $(A+A) \times (A+A)$. We take a subset, S, of this sumset,

$$S = \bigcup_{i=1}^{m-1} (r_i \cap A \times A) + (r_{i+1} \cap A \times A) \subset (A + A) \times (A + A).$$

Simple elementary geometry shows (see the picture below) that the sumsets in the terms are disjoint and each term has $|r_i \cap A \times A||r_{i+1} \cap A \times A|$ elements. Therefore

$$|S| = |AA|\big(|A|^2/|AA|\big)^2 \leqslant |A + A|^2.$$

After rearranging the inequality we get $|A|^4 \leqslant |AA||A + A|^2$, as we wanted. Now we will show a rigorous proof based on this observation.

We are going to use the notation of *multiplicative energy*. The name of this quantity comes from a paper of Tao [13], however its discrete version was used earlier, like in [4].

Let A be a finite set of reals. The multiplicative energy of A, denoted by $E(A)$, is given by

$$E(A) = \big|\big\{(a, b, c, d) \in A^4 \;\big|\; \exists \lambda \in \mathbb{R}\colon\ (a, b) = (\lambda c, \lambda d)\big\}\big|.$$

In the notation of Gowers [8], the quantity $E(A)$ counts the number of quadruples in $\log A$.

To establish the proof of Theorem 2.1 we show the following lemma.

Lemma 2.3. *Let A be a finite set of positive real numbers. Then*

$$\frac{E(A)}{\lceil \log|A| \rceil} \leqslant 4|A + A|^2.$$

Theorem 2.1 follows from Lemma 2.3 via the Cauchy–Schwartz type inequality

$$E(A) \geqslant \frac{|A|^4}{|AA|}. \qquad \square$$

2.2. Proof of Lemma 2.3

Another way of counting $E(A)$ is the following:

$$E(A) = \sum_{x \in A/A} |xA \cap A|^2. \tag{1}$$

The summands on the right hand side can be partitioned into $\lceil \log|A| \rceil$ classes according to the size of $xA \cap A$.

$$E(A) = \sum_{i=0}^{\lceil \log|A| \rceil} \sum_{\substack{x \\ 2^i \leqslant |xA \cap A| < 2^{i+1}}} |xA \cap A|^2.$$

There is an index, I, that

$$\frac{E(A)}{\lceil \log|A| \rceil} \leqslant \sum_{\substack{x \\ 2^I \leqslant |xA \cap A| < 2^{I+1}}} |xA \cap A|^2.$$

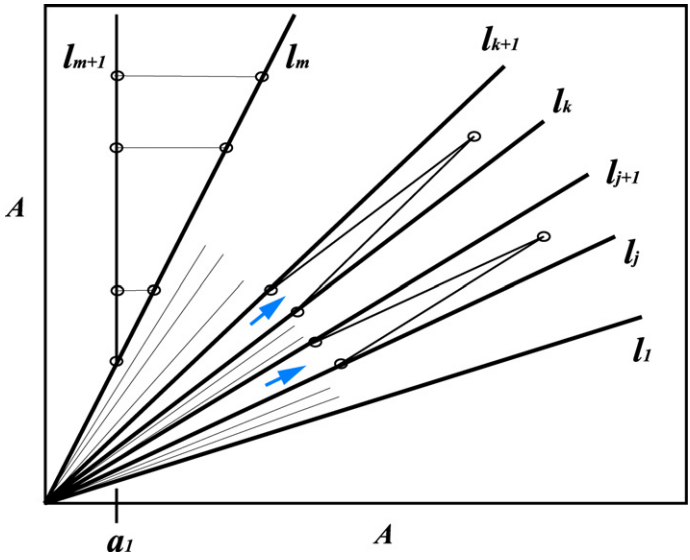


Fig. 1.

Let $D = \{s\colon 2^I \leqslant |sA \cap A| < 2^{I+1}\}$, and let $s_1 < s_2 < \cdots < s_m$ denote the elements of D, labeled in increasing order,

$$\frac{E(A)}{\lceil \log |A| \rceil} \leqslant \sum_{\substack{x \\ 2^I \leqslant |xA \cap A| < 2^{I+1}}} |xA \cap A|^2 < m2^{2I+2}. \tag{2}$$

Each line $l_j\colon y = s_j x$, where $1 \leqslant j \leqslant m$, is incident to at least 2^I and less than 2^{I+1} points of $A \times A$. For easier counting we add an extra line to the set, l_{m+1}, the vertical line through the smallest element of A, denoted by a_1. Line l_{m+1} has $|A|$ points from $A \times A$, however we are considering only the orthogonal projections of the points of l_m. (See Fig. 1.)

The sumset,[1] $(l_i \cap A \times A) + (l_k \cap A \times A)$, $1 \leqslant j < k \leqslant m$, has size $|l_i \cap A \times A||l_k \cap A \times A|$, which is between 2^{2I} and 2^{2I+2}. Also, the sumsets along consecutive line pairs are disjoint, i.e.

$$\big((l_i \cap A \times A) + (l_{i+1} \cap A \times A)\big) \cap \big((l_k \cap A \times A) + (l_{k+1} \cap A \times A)\big) = \emptyset,$$

for any $1 \leqslant j < k \leqslant m$.

The sums are elements of $(A + A) \times (A + A)$, so we have the following inequality,

$$m2^{2I} \leqslant \left| \bigcup_{i=1}^{m} (l_i \cap A \times A) + (l_{i+1} \cap A \times A) \right| \leqslant |A + A|^2.$$

The inequality above with inequality (2) proves the lemma. □

[1] As customary, by the sum of two points on $\mathbb{R}^2$ we mean the point which is the sum of their position vectors.

2.3. Remarks

Let A and B be finite sets of reals. The multiplicative energy, $E(A, B)$, is given by

$$E(A, B) = \left|\left\{(a, b, c, d) \in A \times B \times A \times B \;\middle|\; \exists \lambda \in \mathbb{R}\colon\ (a, b) = (\lambda c, \lambda d)\right\}\right|.$$

In the proof of Lemma 2.3 we did not use the fact that $A = B$, the proof works for the asymmetric case as well. Suppose that $|A| \geqslant |B|$. With the lower bound on the multiplicative energy

$$E(A, B) \geqslant \frac{|A|^2|B|^2}{|AB|}$$

our proof gives the more general inequality

$$\frac{|A|^2|B|^2}{|AB|} \leqslant 4\lceil \log|B| \rceil |A + A||B + B|.$$

3. Very small productsets

In this section we extend our method from two to higher dimensions. We are going to consider lines though the origin as before, however there is no notion of consecutiveness among these lines in higher dimensions available. We will consider them as points in the projective real space and will find a triangulation of the pointset. The simplices of the triangulation will define the neighbors among the selected lines.

The sum-product bound in Theorem 2.1 is asymmetric. It shows that the productset should be very large if the sumset is small. On the other hand it says almost nothing in the range where the productset is small. For integers, Chang [2] proved that there is a function $\delta(\varepsilon)$ that if $|AA| \leqslant |A|^{1+\varepsilon}$ then $|A + A| \geqslant |A|^{2-\delta}$, where $\delta \to 0$ if $\varepsilon \to 0$. A similar result is not known for reals. It follows from Elekes' bound [3] (and also from Theorem 2.1) that there is a function $\delta(\varepsilon)$ that if $|AA| \leqslant |A|^{1+\varepsilon}$ then $|A + A| \geqslant |A|^{3/2-\delta}$, where $\delta \to 0$ if $\varepsilon \to 0$. We prove here a generalization of this bound for k-fold sumsets. For any integer $k \geqslant 2$ the k-fold subset of A, denoted by kA is the set

$$kA = \{a_1 + a_2 + \cdots + a_k \mid a_1, \ldots, a_k \in A\}.$$

Theorem 3.1. *For any integer $k \geqslant 2$ there is a function $\delta = \delta_k(\varepsilon)$ that if $|AA| \leqslant |A|^{1+\varepsilon}$ then $|kA| \geqslant |A|^{2-1/k-\delta}$, where $\delta \to 0$ if $\varepsilon \to 0$.*

Proof. We can suppose that A has only positive elements WLOG. Let $|AA| \leqslant |A|^{1+\varepsilon}$. By a Plünnecke type inequality (Corollary 5.2 [11] or Chapter 6.5 [14]) we have $|A/A| \leqslant |A|^{1+2\varepsilon}$. Consider the k-fold Cartesian product $A \times A \times \cdots \times A$, denoted by $\times^k A$. It can be covered by no more than $|A/A|^{k-1}$ lines going through the origin. Fig. 2 illustrates the $k = 3$ case. Let H denote the set of lines through the origin containing at least $|A|^{1-2\varepsilon(k-1)}/2$ points of $\times^k A$. With this selection, the lines in H cover at least half of the points in $\times^k A$ since

$$\frac{|A|^{1-2\varepsilon(k-1)}}{2}|A/A|^{k-1} = \frac{|A|^k}{2|A|^{(1+2\varepsilon)(k-1)}}|A/A|^{k-1} \leqslant \frac{|A|^k}{2}.$$

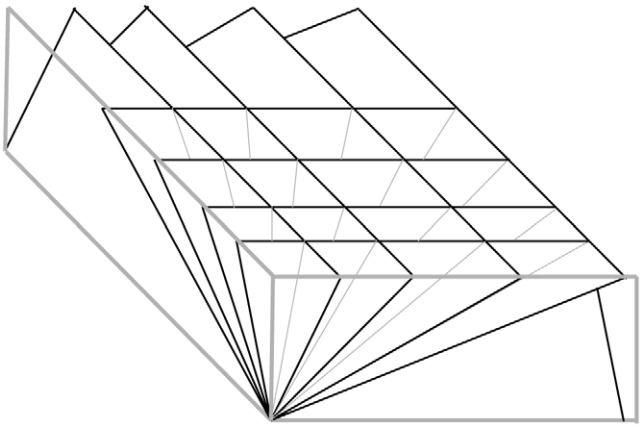


Fig. 2.

As no line has more than $|A|$ points common with $\times^k A$, therefore $|H| \geqslant |A|^{k-1}/2$. The set of lines, H, represents a set of points, P, in the projective real space $\mathbb{R}P^{k-1}$. Point set P has full dimension $k-1$ as it has a nice symmetry. The symmetry follows from the Cartesian product structure; if a point with coordinates $(a_1, \ldots, a_k)$ is in P then the point $(\sigma(a_1), \ldots, \sigma(a_k))$ is also in P for any permutation $\sigma \in S_k$. Let us triangulate P. By triangulation we mean a decomposition of the convex hull of P into non-degenarate, $k-1$-dimensional, simplices such that the intersection of any two is the empty set or a face of both simplices and the vertex set of the triangulation is P. It is not obvious that such triangulation always exists. For the proof we refer to Chapter 7 in [7] or Chapter 2 in [9]. The size of the triangulation (the number of simplices in the triangulation) is at least $|P| - (k-1)$. It is possible that for sets with symmetries like P the maximum triangulation size is much larger, however we were unable to find a better bound. For similar problems about maximum triangulations see [1]. Let $\tau(P)$ be a triangulation of P. We say that k lines $l_1, \ldots, l_k \in H$ form a simplex if the corresponding points in P are vertices of a simplex of the triangulation. We use the following notation for this: $\{l_1, \ldots, l_k\} \in \tau(P)$. In the two-dimensional case we used that the sumsets of points on consecutive lines are disjoint. Here we are using that the interiors of the simplices are disjoint, therefore sumsets of lines of simplices are also disjoint. Note that we assumed that A is positive, so we are considering convex combinations of vectors with positive coefficients. Let $\{l_1, \ldots, l_k\} \in \tau(P)$ and $\{l'_1, \ldots, l'_k\} \in \tau(P)$ are two distinct simplices. Then

$$\left(\sum_{i=1}^{k} l_i \cap \times^k A\right) \cap \left(\sum_{i=1}^{k} l'_i \cap \times^k A\right) = \emptyset.$$

Also, since the k vectors parallel to the lines $\{l_1, \ldots, l_k\} \in \tau(P)$ are linearly independent, all sums are distinct,

$$\left|\sum_{i=1}^{k} l_i \cap \times^k A\right| = \prod_{i=1}^{k} |l_i \cap \times^k A|.$$

Now we are ready to put everything together into a sequence of inequalities proving Theorem 3.1,

$$|kA|^k \geqslant \sum_{\{l_1,\ldots,l_k\}\in\tau(P)} \left|\sum_{i=1}^{k} l_i \cap \times^k A\right| \geqslant \left(|A|^{k-1} - k + 1\right) \prod_{i=1}^{k} \left|l_i \cap \times^k A\right|.$$

Every line is incident to at least $|A|^{1-2\varepsilon(k-1)}/2$ points of $\times^k A$, therefore

$$|kA|^k \geqslant \frac{|A|^{k-1+k(1-2\varepsilon(k-1))} - (k-1)|A|^{k(1-2\varepsilon(k-1))}}{2^k}.$$

Taking the kth root of both sides we get the result we wanted to show

$$|kA| \geqslant c_k |A|^{2-1/k-2(k-1)\varepsilon}. \qquad \square$$

References

[1] P. Brass, On the size of higher dimensional triangulations, in: J.E. Goodman, et al. (Eds.), Combinatorial and Computational Geometry, in: Math. Sci. Res. Inst. Publ., vol. 52, Cambridge University Press, 2005, pp. 145–151.
[2] M.C. Chang, The Erdős–Szemerédi problem on sum set and product set, Ann. of Math. 157 (2003) 939–957.
[3] Gy. Elekes, On the number of sums and products, Acta Arith. 81 (1997) 365–367.
[4] Gy. Elekes, Sums versus products in number theory, algebra and Erdős geometry—A survey, in: Paul Erdős and His Mathematics II, in: János Bolyai Math. Soc. Stud., vol. 11, János Bolyai Math. Soc., Budapest, 2002, pp. 241–290.
[5] P. Erdős, E. Szemerédi, On sums and products of integers, in: Stud. Pure Math., Birkhäuser, Basel, 1983, pp. 213–218.
[6] K. Ford, Sums and products from a finite set of real numbers, Ramanujan J. 2 (1998) 59–66.
[7] I.M. Gelfand, M.M. Kapranov, A.V. Zelevinsky, Discriminants, Resultants, and Multidimensional Determinants, Modern Birkhäuser Classics, Birkhäuser, Boston, 2008, reprint of the 1994 edition, 512 pp.
[8] W.T. Gowers, A new proof of Szemerédi's theorem for arithmetic progressions of length four, Geom. Funct. Anal. 8 (1998) 529–551.
[9] J.A. De Lorea, J. Rambau, F.S. Leal, Triangulations: Structures and Algorithms, Springer, forthcoming.
[10] M.B. Nathanson, On sums and products of integers, Proc. Amer. Math. Soc. 125 (1997) 9–16.
[11] I.Z. Ruzsa, An application of graph theory to additive number theory, Scientia 3 (1989) 97–109.
[12] J. Solymosi, On the number of sums and products, Bull. London Math. Soc. 37 (4) (2005) 491–494.
[13] T. Tao, Product set estimates for non-commutative groups, Combinatorica, DOI 10.1007/s00493-008-2271-7, arXiv:math.CO/0601431.
[14] T. Tao, V. Vu, Additive Combinatorics, Cambridge Stud. Adv. Math., vol. 105, Cambridge University Press, 2006.

SIAM J. DISCRETE MATH.
Vol. 0, No. 0, pp. 000–000

DENSE ARRANGEMENTS ARE LOCALLY VERY DENSE. I*

JÓZSEF SOLYMOSI†

Abstract. The Szemerédi–Trotter theorem [*Combinatorica*, 3 (1983), pp. 381–392] gives a bound on the maximum number of incidences between points and lines on the Euclidean plane. In particular it says that n lines and n points determine $O(n^{4/3})$ incidences. Let us suppose that an arrangement of n lines and n points defines $cn^{4/3}$ incidences, for a given positive c. It is widely believed that such arrangements have special structure, but no results are known in this direction. Here we show that for any natural number, k, one can find k points of the arrangement in general position such that any pair of them is incident to a line from the arrangement, provided by $n \geq n_0(k)$. In a subsequent paper we will a establish a similar statement for hyperplanes.

Key words. point-line incidences, Szemerédi–Trotter theorem, regularity lemma

AMS subject classifications. 52C10, 52C30, 52C45

DOI. 10.1137/05062826X

1. Introduction.

The celebrated Szemerédi–Trotter theorem [21] states that for n points on the plane, the number of m-rich lines cannot exceed

$$O(n^2/m^3 + n/m), \tag{1.1}$$

and this bound is tight in the worst case. This result has numerous applications not only in geometry [11, 22], but also in number theory [4]. The Szemerédi–Trotter theorem has various proofs; the most elegant is Székely's [22]. However, the proofs provide very limited insight view of the structure of extremal arrangements. It is widely believed that a point-line arrangement which defines many incidences has a special, somehow rigid structure. For example, let us mention here a question of Elekes. Is it true that for every $c > 0$ there is a $c' > 0$ such that if a set of n points on the plane contains at least cn^2 collinear triples, then at least $n^{c'}$ points are along an algebraic curve of degree d, where d is a universal constant?

The main purpose of this paper is to show that any arrangement with close to the maximum number of incidences is locally a collection of complete geometric graphs. For the sake of simplicity we state the theorem for the balanced case, when the number of lines equals the number of points, but it is quite straightforward to see the similar statement for unbalanced cases as well.

Recent work of Gowers [6] and Nagle, Rödl, Schacht, and Skokan [9, 12, 13] has established a hypergraph removal lemma, which in turn implies similar results to hyperplanes; however, a slightly different approach is needed, mainly because the higher dimensional extensions of the Szemerédi–Trotter theorem are not as well defined as in the planar case. To obtain sharp bounds one needs certain restrictions on the arrangements. Therefore the corresponding structure theorems will appear in a subsequent paper.

*Received by the editors April 1, 2005; accepted for publication (in revised form) February 27, 2006; published electronically DATE. This research was supported by OTKA and NSERC grants.
http://www.siam.org/journals/sidma/x-x/62826.html

†Department of Mathematics, University of British Columbia, Vancouver, BC, Canada V6T 1Z2 (solymosi@math.ubc.ca).

A point set or a set of lines is in *general position* if no three of the elements are collinear or concurrent.

THEOREM 1.1. *For every natural number k and real $c > 0$ there is a threshold $n_0 = n_0(k, c)$ such that if an arrangement of $n \geq n_0$ lines and n points defines at least $cn^{4/3}$ incidences, then one can always find k points of the arrangement in general position, such that any pair of them is incident to a line from the arrangement.*

As we will see from the proof, the complete k-tuple is "local" in the sense that for any pair of points of the k-tuple, p_1 and p_2, the number of points from the arrangement, incident to the line segment (p_1, p_2), is less than k.

2. Proof of Theorem 1.1. The main tool of the proof is Szemerédi's regularity lemma [19, 20]. We will use its *counting lemma* form, because it is easier to extend to hypergraphs which we will need for the higher dimensional extensions. Let us prove first the simplest case, to show that there is always a triangle. This "simplest case" is interesting in its own right; the statement of Lemma 2.1 implies Roth's theorem [14] about arithmetic progressions on dense subsets of integers. For the details we refer to [16, 17].

LEMMA 2.1. *For every $c > 0$ there is a threshold $n_0 = n_0(c)$ and a positive $\delta = \delta(c)$ such that, for any set of $n \geq n_0$ lines L and any set of $m \geq cn^2$ points P, if every point is incident to three lines, then there are at least δn^3 triangles in the arrangement. (A triangle is a set of three distinct points from P such that any two are incident to a line from L.)*

This lemma follows the following theorem of Ruzsa and Szemerédi [15], which is also called the *triangle removal lemma* or the counting lemma for triangles.

THEOREM 2.2 (see [15]). *Let G be a graph on n vertices. If G is the union of cn^2 edge-disjoint triangles, then G contains at least δn^3 triangles, where δ depends on c only.*

The same theorem from a different angle is the following.

THEOREM 2.3. *Let G be a graph on n vertices. If G contains $o(n^3)$ triangles, then one can remove $o(n^2)$ edges to make G triangle-free.*

To prove Lemma 2.1, let us construct a graph where L is the vertex set and two vertices are adjacent if and only if the corresponding lines cross at a point of P. This graph is the union of cn^2 disjoint triangles, every point of P defines a unique triangle, so we can apply Theorem 2.2.

To determine the number of triangles in any arrangement of lines and points seems to be a hard task. A related conjecture of de Caen and Székely [1] is that n points and m lines cannot determine more than nm triangles.

One can repeat the same argument, now with k instead of 3. The corresponding counting lemma can be proven using Szemerédi's regularity lemma. The proof is analogous to the Ruzsa–Szemerédi theorem. There are slightly different ways to state the regularity lemma; for our purposes the so called *degree form* is convenient. For the notations and proofs we refer to the survey paper of Komlós and Simonovits [7].

THEOREM 2.4 (regularity lemma). *For every $\epsilon > 0$ there is an $M = M(\epsilon)$ such that if $G = (V, E)$ is any graph and $d \in (0, 1]$ is any real number, then there is a partition of the vertex set V into $k + 1$ clusters $V_0, V_1, \ldots, V_k$, and there is a subgraph $G' \subset G$ with the following properties:*

- *$k \leq M$,*
- *$|V_0| \leq \epsilon |V|$,*
- *all clusters V_i, $i \geq 1$, are of the same size $m \leq \lceil \epsilon |V| \rceil$,*
- *$\deg_{G'}(v) > \deg_G(v) - (d + \epsilon)|V|$ for all $v \in V$,*

- $e(G'(V_i)) = 0$ *for each* $i \geq 1$,
- *all pairs* $G'(V_i, V_j)$ $(1 \leq i < j \leq k)$ *are* ϵ*-regular, each with a density either* 0 *or greater than* d.

Armed with the regularity lemma we are ready to prove the following statement, which is crucial in the proof of Theorem 1.1.

LEMMA 2.5. *For every* $c > 0$ *there is a threshold* $n_0 = n_0(c)$ *and a positive* $\delta = \delta(c)$ *such that, for any set of* $n \geq n_0$ *lines* L *and any set of* $m \geq cn^2$ *points* P*, if every point is incident to* k *lines, then there are at least* δn^k *complete* k*-tuples in the arrangement. (A complete* k*-tuple is a set of* k *distinct lines in general position from* L *such that any two intersect in a point from* P*.)*

Proof. To avoid having too many degenerate k-tuples, we remove some points from P which have many lines incident to them. Let P', which is the subset of P, consist of points incident to at most $100/c$ lines from L. We can apply (1.1) to see that P' is a large subset of P, say $2|P'| > |P|$. Let us construct a graph G where L is the vertex set and two vertices are adjacent if and only if the corresponding lines cross at a point of P'. This graph, G, is the union of at least $\frac{c}{2}n^2$ edge-disjoint K_k−s. Find a subgraph, G', provided by Theorem 2.4 with $\epsilon \ll c$. In G' we still have some complete K_k−s (when going from G to G' we removed $(\epsilon + d)n^2$ edges, much less than cn^2). The edges of such a complete graph are connecting V_i−s such that the bipartite graphs between them are dense and regular. This already implies the existence of many complete subgraphs, K_k−s, as the following lemma, quoted from [7], shows. □

LEMMA 2.6. *Given* $d > \epsilon > 0$*, a graph* R *on* k *vertices, and a positive integer* m*, let us construct a graph* G *by replacing every vertex of* R *by* m *vertices, and replacing the edges of* R *with* ϵ*-regular pairs of density at least* d*. Then* G *has at least* αm^k *copies of* R*, where* α *depends on* ϵ, d*, and* k*, but not on* m.

Most of the complete k-vertex subgraphs of graph G' define a complete k-tuple in the arrangement, i.e., the corresponding lines are in general position. To see this, let us count the "degenerate" k-tuples, where at least one triple is concurrent. The number of concurrent triples is at most $cn^2\binom{100/c}{3} \leq c'n^2$. For every concurrent triple one can select $k-3$ lines to get a degenerate k-tuple. The expression $c'n^{k-1}$ is clearly an upper bound on the degenerate k-tuples; therefore most of the complete graphs on k vertices in G' are complete k-tuples if n is large enough, $n \geq n_0 = n_0(c)$.

The final step of the proof of Theorem 1.1 is to show that arrangements with many incidences always have a substructure where one uses Lemma 2.1. We divide the arrangement into smaller parts where we apply the dual of Lemma 2.1. The common technique to do that is so-called *cutting*, which was introduced by Chazelle (see in [2] or in [10]) about 20 years ago. Here we use a more general result, a theorem of Matousek [8].

LEMMA 2.7. *Let* P *be a point set,* $P \subset \boldsymbol{R}^d, |P| = n$*, and let* r *be a parameter,* $1 \ll r \ll n$*. Then the set* P *can be partitioned into* t *sets* $\Delta_1, \Delta_2, \ldots, \Delta_t$*, in such a way that* $n/r \leq |\Delta_i| \leq 2n/r$ *for all* i*, and any hyperplane crosses no more than* $O(r^{1-1/d})$ *sets, where* $t = O(r)$.

One can use the $d = 2$ case and we choose the value $r = \beta_k n^{2/3}$, where β_k is a constant that depends on k and which we will specify later. Let us count the number of incidences along the lines of L, according to the partition of P. For a given line $\xi \in L$, we count the sum $\sum_{i=1}^{t} \lfloor |\Delta_i \bigcap \xi|/k \rfloor$, which is not much smaller than the number of incidences on ξ over k if ξ is rich of incidences, say, incident to much more than $r^{1/2}k$ points of P. From the condition of Theorem 1.1 and the properties of the

partition we have the following inequality:

$$\frac{c}{k}n^{4/3} \leq \sum_{\xi\in L}\sum_{i=1}^{t}\left\lfloor \frac{|\Delta_i \bigcap \xi|}{k}\right\rfloor + |L|r^{1/2}.$$

Choosing $\beta_k = \frac{c}{2k}$, the inequality becomes

$$\frac{cn^{4/3}}{2k} = c_k n^{\frac{4}{3}} \leq \sum_{\xi\in L}\sum_{i=1}^{t}\left\lfloor \frac{|\Delta_i \bigcap \xi|}{k}\right\rfloor = \sum_{i=1}^{t}\sum_{\xi\in L}\left\lfloor \frac{|\Delta_i \bigcap \xi|}{k}\right\rfloor.$$

Therefore there is an index i, such that

$$c_k n^{2/3} \leq \sum_{\xi\in L}\left\lfloor \frac{|\Delta_i \bigcap \xi|}{k}\right\rfloor.$$

If $s = \lfloor \frac{|\Delta_i \cap \xi|}{k} \rfloor$, then we can partition the points incident to ξ into s consecutive k-tuples. We can break the line into s k-rich line segments and consider them as separate lines. Our combinatorial argument in Lemma 2.5 is robust enough to allow such modifications. Then we have some $c'n^{2/3}$ k-rich lines on $|\Delta_i| = c''n^{1/3}$ points. (Another possible way to show that there are at least $c'n^{2/3}$ k-rich lines is to apply the Szemerédi–Trotter theorem, (1.1), to show that most of the lines are not "very rich.") To complete the proof of Theorem 1.1, we apply the dual statement of Lemma 2.5.

REFERENCES

[1] D. de Caen and L. A. Székely, *On dense bipartite graphs of girth eight and upper bounds for certain configurations in planar point-line systems*, J. Combin. Theory Ser. A., 77 (1997), pp. 268–278.
[2] B. Chazelle, *The Discrepancy Method*, Cambridge University Press, Cambridge, UK, 2000.
[3] K. L. Clarkson, H. Edelsbrunner, L. J. Guibas, M. Sharir, and E. Welzl, *Combinatorial complexity bounds for arrangements of curves and spheres*, Discrete Comput. Geom., 5 (1990), pp. 99–160.
[4] Gy. Elekes, *SUMS versus PRODUCTS in number theory, algebra and Erdős geometry*, in Paul Erdos and His Mathematics II, Bolyai Soc. Math. Stud. 11, János Bolyai Math. Soc., Budapest, 2002, pp. 241–290.
[5] Gy. Elekes and Cs. D. Tóth, *Incidences of not-too-degenerate hyperplanes*, in Proceedings of the 21st ACM Symposium in Computer Geometrics (Pisa, 2005), ACM Press, New York, pp. 16–21.
[6] W. T. Gowers, *Hypergraph Regularity and the Multidimensional Szemerédi Theorem*, preprint.
[7] J. Komlós and M. Simonovits, *Szemerédi's regularity lemma and its applications in graph theory*, in Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993), Bolyai Soc. Math. Stud. 2, János Bolyai Math. Soc., Budapest, 1996, pp. 295–352.
[8] J. Matoušek, *Efficient partition trees*, Discrete Comput. Geom. 8 (1992), pp. 315–334.
[9] B. Nagle, V. Rödl, and M. Schacht, *The counting lemma for regular k-uniform hypergraphs*, Random Structures Algorithms, 28 (2006), pp. 113–179.
[10] J. Pach and P. K. Agarwal, *Combinatorial Geometry*, John Wiley, New York, 1995.
[11] J. Pach and M. Sharir, *Geometric incidences*, in Towards a Theory of Geometric Graphs, Contemp. Math. 342, AMS, Providence, RI, 2004, pp. 185–223.
[12] V. Rödl and J. Skokan, *Regularity lemma for k-uniform hypergraphs*, Random Structures Algorithms, 25 (2004), pp. 1–42.
[13] V. Rödl and J. Skokan, *Applications of the regularity lemma for uniform hypergraphs*, Random Structures Algorithms, 28 (2006), pp. 180–194.
[14] K. F. Roth, *On certain sets of integers*, J. London Math. Soc., 28 (1953), pp. 245–252.
[15] I. Ruzsa and E. Szemerédi, *Triple systems with no six points carrying three triangles*, Colloq. Math. Soc. Janos Bolyai, 18 (1978), pp. 939–945.

[16] J. Solymosi, *Note on a generalization of Roth's theorem*, in Discrete and Computational Geometry, Algorithms Combin. 25, Springer-Verlag, Berlin, 2003, pp. 825–827.

[17] J. Solymosi, *A note on a question of Erdős and Graham*, Combin. Probab. Comput., 13 (2004), pp. 263–267.

[18] J. Solymosi and Cs. D. Tóth, *Distinct distances in the plane*, Discrete Comput. Geom., 25 (2001), pp. 629–634.

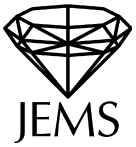
[19] E. Szemerédi, *On sets of integers containing no four elements in arithmetic progression*, Acta Math. Acad. Sci. Hungar., 20 (1969), pp. 89–104.

[20] E. Szemerédi, *Regular partitions of graphs*, in Problèmes Combinatoires et Théorie des Graphes, Proc. Colloque Inter. CNRS, Bermond et al., eds., CNRS, Paris, 1978, pp. 399–401.

[21] E. Szemerédi and W. T. Trotter Jr., *Extremal problems in discrete geometry*, Combinatorica, 3 (1983), pp. 381–392.

[22] L. A. Székely, *Crossing numbers and hard Erdős problems in discrete geometry*, Combin. Probab. Comput., 6 (1997), pp. 353–358.

J. Eur. Math. Soc. 9, 1–16



József Solymosi and Mei-Chu Chang

Sum-product theorems and incidence geometry

Received June 8, 2004, and in revised form March 30, 2006

Abstract. We prove the following theorems in incidence geometry.

1. There is $\delta > 0$ such that for any $P_1, \dots, P_4 \in \mathbb{C}^2$, and $Q_1, \dots, Q_n \in \mathbb{C}^2$, if there are $\leq n^{(1+\delta)/2}$ distinct lines between P_i and Q_j for all i, j, then $P_1, \dots, P_4$ are collinear. If the number of the distinct lines is $< cn^{1/2}$, then the cross ratio of the four points is algebraic.
2. Given $c > 0$, there is $\delta > 0$ such that for any $P_1, P_2, P_3 \in \mathbb{C}^2$ noncollinear, and $Q_1, \dots, Q_n \in \mathbb{C}^2$, if there are $\leq cn^{1/2}$ distinct lines between P_i and Q_j for all i, j, then for any $P \in \mathbb{C}^2 \setminus \{P_1, P_2, P_3\}$, we have δn distinct lines between P and Q_j.
3. Given $c > 0$, there is $\epsilon > 0$ such that for any $P_1, P_2, P_3 \in \mathbb{C}^2$ (respectively, $\mathbb{R}^2$) collinear, and $Q_1, \dots, Q_n \in \mathbb{C}^2$ (respectively, $\mathbb{R}^2$), if there are $\leq cn^{1/2}$ distinct lines between P_i and Q_j for all i, j, then for any P not lying on the line $L(P_1, P_2)$, we have at least $n^{1-\epsilon}$ (resp. $n/\log n$) distinct lines between P and Q_j.

The main ingredients used are the subspace theorem, Balog–Szemerédi–Gowers theorem, and Szemerédi–Trotter theorem. We also generalize the theorems to higher dimensions, extend Theorem 1 to $\mathbb{F}_p^2$, and give the version of Theorem 2 over $\mathbb{Q}$.

0. Introduction

Notation.

- For $P \neq Q$, $L(P, Q)$ denotes the line through P, Q.
- Let A be a subset of a ring. Then $2A = \{a + a' : a, a' \in A\}$, $A^2 = \{aa' : a, a' \in A\}$.

We first prove the following two theorems.

Theorem 1. *There is $\delta > 0$ such that for any $P_1, \dots, P_4 \in \mathbb{C}^2$, and $Q_1, \dots, Q_n \in \mathbb{C}^2$, if*

$$|\{L(P_i, Q_j) : 1 \leq i \leq 4,\ 1 \leq j \leq n\}| \leq n^{(1+\delta)/2}, \tag{0.1}$$

then $P_1, \dots, P_4$ are collinear. If

$$|\{L(P_i, Q_j) : 1 \leq i \leq 4,\ 1 \leq j \leq n\}| \leq cn^{1/2}, \tag{0.2}$$

then the cross ratio of $P_1, \dots, P_4$ is algebraic.

J. Solymosi: Mathematics Department, University of British Columbia, Vancouver, BC V6T 1Z2, Canada; e-mail: solymosi@math.ubc.ca

M.-C. Chang: Mathematics Department, University of California, Riverside, CA 92521, USA; e-mail: mcc@math.ucr.edu

Theorem 2. *Given $c > 0$, there is $\delta > 0$ such that for any $P_1, P_2, P_3 \in \mathbb{C}^2$ noncollinear, and $Q_1, \dots, Q_n \in \mathbb{C}^2$, if*

$$|\{L(P_i, Q_j) : 1 \le i \le 3,\ 1 \le j \le n\}| \le cn^{1/2}, \tag{0.3}$$

then for any $P \in \mathbb{C}^2 \setminus \{P_1, P_2, P_3\}$, we have

$$|\{L(P, Q_j) : 1 \le j \le n\}| = \delta n. \tag{0.4}$$

Theorem 3. *Given $c > 0$, there is $\epsilon > 0$ such that for any $P_1, P_2, P_3 \in \mathbb{C}^2$ collinear, and $Q_1, \dots, Q_n \in \mathbb{C}^2$, if*

$$|\{L(P_i, Q_j) : 1 \le i \le 3,\ 1 \le j \le n\}| \le cn^{1/2}, \tag{0.5}$$

then for any $P \in \mathbb{C}^2 \setminus L(P_1, P_2)$, we have

$$|\{L(P, Q_j) : 1 \le j \le n\}| > n^{1-\epsilon}. \tag{0.6}$$

Remark 4. In Theorem 3, the bound $n^{1-\epsilon}$ in (0.6) is replaced by $n/\log n$ if the points are in $\mathbb{R}^2$ instead of $\mathbb{C}^2$.

Remark 5. In Remark 1.1 below, we see that assumption (0.3) does occur.

We will first interpret the geometric problems under consideration as sum-product problems. Roughly speaking, for Theorem 2, we want to show that given two sets $C, D \subset \mathbb{C}^2$ of about the same size, if $\{d_i/c_i : (c_i, d_i) \in C \times D,\ 1 \le i \le n\}$ is small, then $\{(d_i + b)/(c_i + a) : (c_i, d_i) \in C \times D,\ 1 \le i \le n\}$ is large, where a, b are fixed. So we want to have an upper bound on the number of solutions (c_i, d_i, c_j, d_j) of the equation

$$\frac{d_i + b}{c_i + a} = \frac{d_j + b}{c_j + a}.$$

This interpretation is introduced in Section 1. In Section 2, we use the subspace theorem to prove Theorem 2, for the case when the point P is not on any line connecting the P_i's. In Section 3, we use the Szemerédi–Trotter theorem to prove the corresponding case of Theorem 1. We also give a short proof using a theorem about convex functions by Elekes, Nathanson and Ruzsa [ENR]. The argument using the Szemerédi–Trotter theorem [S], besides applying over $\mathbb{C}$ (rather than $\mathbb{R}$), has the advantage that the set-up (reducing the problem to bounding the number of solutions of equations) was already used for the subspace theorem approach. Also, it generalizes easily to the prime field $\mathbb{F}_p$ setting. In Section 4, we use the sum-product theorem to take care of all the cases when more than two of the P_i's are at infinity. In Section 5, we generalize the theorems to high dimensions. In Section 6, we prove a stronger theorem over $\mathbb{Q}$ by using the λ_q constant (see [BC]).

This work is one more illustration of the relations between arithmetic combinatorics and point-line incidence geometry. Let us recall that presently the strongest results on the sum-product problem were obtained using the Szemerédi–Trotter theorem (due to Elekes and the second author). The results in this paper are another demonstration of the interplay between these two fields.

1. The set-up

Our strategy of proving Theorem 1 is to assume that P_1, P_2, P_3 are not collinear and get a large family of lines $L(P_4, Q_j)$ violating assumption (0.1). Therefore, the settings for Theorem 1 and Theorem 2 are the same. For simplicity, we describe the situation for Theorem 2 here and indicate the (small) difference when we prove Theorem 1.

We will work in the projective space $\mathbb{CP}^2 \cong (\mathbb{C}^3 \setminus \{0\})/\sim$, where $(x, y, z) \sim (\lambda x, \lambda y, \lambda z)$ for any $\lambda \neq 0$. We identify $\mathbb{C}^2$ with the affine space in $\mathbb{CP}^2$ defined by $z \neq 0$ via $(x, y) \mapsto (x, y, 1)$.

Let L_∞ be the line at infinity defined by $z = 0$. We may assume

(i) P_1, P_2, P_3 are $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$. (Clearly, P_1 and P_2 lie on L_∞.)
(ii) No Q_i lies on L_∞.

In fact, let A be the 3×3 matrix with the vector P_i as the ith column. Since the P_i's are not collinear, the matrix A is invertible. Hence the linear transformation $T : \mathbb{C}^3 \to \mathbb{C}^3$ defined by $P \mapsto A^{-1}P^T$ sends P_1, P_2, P_3 to $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$. To see (ii), we notice that for any $Q = (1, d, 0) \in L_\infty$, the line $L(Q, P_3)$ is defined by $y = dx$. Assumption (0.3) implies that $|\{Q_i : Q_i \in L_\infty\}| \leq cn^{1/2} \ll n$.

Let

$$Q_i = (c_i, d_i, 1),$$
$$C = \{c_i : 1 \leq i \leq n\}, \quad D = \{d_i : 1 \leq i \leq n\} \tag{1.1}$$
$$\mathcal{G} = \{(c_i, d_i) : 1 \leq i \leq n\}, \quad C^{-1} \underset{\mathcal{G}}{\times} D = \{d_i/c_i : 1 \leq i \leq n\}. \tag{1.2}$$

Then

$$|\mathcal{G}| = n \tag{1.3}$$

and assumption (0.3) implies

$$|C^{-1} \underset{\mathcal{G}}{\times} D| \leq cn^{1/2}, \quad |C| = |D| = c'n^{1/2}, \tag{1.4}$$

since the lines $L(P_1, Q_i)$, $L(P_2, Q_i)$, $L(P_3, Q_i)$ are defined by $y = d_i z$, $x = c_i z$, $y = (d_i/c_i)x$, and $|C|\,|D| \geq n$.

Remark 1.1. Assumption (0.3) does occur. For example, if we let $Q_{i,j} = (2^i, 2^j, 1)$, $1 \leq i, j \leq N$, then

$$|\{L(P_1, Q_{i,j})\}_{i,j}| = |\{L(P_2, Q_{i,j})\}_{i,j}| = N, \quad |\{L(P_3, Q_{i,j})\}_{i,j}| = 2N - 1.$$

To be able to apply the tools from sum-product theory, we need the Laczkovich–Ruzsa version [LR] of the Balog–Szemerédi–Gowers theorem.

Theorem BSG-LR. *Let A, B be subsets of an abelian group with $|A| = |B| = N$, and let $G \subset A \times B$ with $|G| > K^{-1}N^2$. Define*

$$A \overset{G}{+} B = \{a + b : (a, b) \in G\}. \tag{1.5}$$

If $|A \overset{G}{+} B| < KN$, *then there are subsets* $A' \subset A$ *and* $B' \subset B$ *such that*

$$|A' + B'| < K^c N$$

and

$$|A'|, |B'| > K^{-c} N. \tag{1.6}$$

Remark 1.2. The absolute constant c in the above theorem is at most 8 (see [SSV]).

2. The proof of Theorem 2 for finite points

Let $N = n^{1/2}$. Take a point $P = (-a, -b, 1) \in \mathbb{C}^2$. The line $L(P, Q_i)$ has slope $(d_i + b)/(c_i + a)$. With the help of Theorem BSG-LR, Theorem 2 is reduced to the following

Theorem 2.1. *Let* $X = \{x_i \in \mathbb{C}^2 : 1 \le i \le N^2\}$ *and* $Y = \{y_i \in \mathbb{C}^2 : 1 \le i \le N^2\}$ *with* $|Y/X| \le cN$ *and* $|X| = |Y| = c'N$. *Fix* $a, b \in \mathbb{C}$. *Define*

$$Z = \left\{ \frac{y_i + b}{x_i + a} : 1 \le i \le N^2 \right\}.$$

Then $|Z| > \delta N^2$ *for some* $\delta > 0$.

Proof. Let $I_z = \{i : (y_i + b)/(x_i + a) = z\}$. Then $\sum_{z \in Z} |I_z| = n = N^2$ and Cauchy–Schwarz gives

$$N^4 \le |Z| \sum |I_z|^2.$$

Now

$$\begin{aligned} \sum |I_z|^2 &= \left| \left\{ (i, j) : \frac{y_i + b}{x_i + a} = \frac{y_j + b}{x_j + a},\ 1 \le i, j \le n \right\} \right| \\ &\le \left| \left\{ (x, x', y, y') \in X \times X \times Y \times Y : \frac{y + b}{x + a} = \frac{y' + b}{x' + a} \right\} \right| \\ &= |\{(x, x', y, y') \in X \times X \times Y \times Y : x'y + bx' + ay = xy' + bx + ay'\}|. \end{aligned} \tag{2.1}$$

To bound (2.1), we invoke the subspace theorem [ESS], which gives an upper bound on the number of solutions of a linear equation in a multiplicative group.

A solution $(x_1, \ldots, x_m)$ of the equation

$$\sum_{i=1}^{m} c_i x_i = 1, \qquad c_i \in \mathbb{C}, \tag{2.2}$$

is called *nondegenerate* if $\sum_{j=1}^{k} c_{i_j} x_{i_j} \ne 0$ for all k. The bound given below is due to Evertse, Schlickewei and Schmidt [ESS].

Subspace Theorem. *Let $\Gamma < \langle \mathbb{C}^*, \cdot \rangle$ be a subgroup of the multiplicative group of $\mathbb{C}$, and let the rank of Γ be r. Then*

$$\left| \left\{ \textit{nondegenerate solutions of } \sum_{i=1}^{m} c_i x_i = 1 \textit{ in } \Gamma \right\} \right| < e^{(r+1)(6m)^{3m}}.$$

The formulation of the subspace theorem we need is the following (see [C2])

Corollary 2.2 ([C2])**.** *Let $\Gamma < \langle \mathbb{C}^*, \cdot \rangle$ be a subgroup of rank r and $A \subset \Gamma$ with $|A| = N$. Then the number of solutions in A of*

$$x_1 + \cdots + x_{2h} = 0 \tag{2.3}$$

is bounded by $N^{h-1}e^{rc} + N^h$, up to a constant depending on h. Here $c = c(h)$.

In order to apply the subspace theorem, we need the following (see [Fr], [R1], [Bi]).

Freiman's Lemma. *Let $\langle G, \cdot \rangle$ be a torsion-free abelian group and $A \subset G$ with $|A^2| < K|A|$. Then*

$$A \subset \{g_1^{j_1} \cdots g_d^{j_d} : j_i = 1, \ldots, \ell_i, \textit{ and } g_i \in G\}, \tag{2.4}$$

where $d \leq K$ and $\prod \ell_i < c(K)|A|$.

We let $\Gamma < \langle \mathbb{C}^*, \cdot \rangle$ be the subgroup generated by $g_1, \ldots, g_d$. Then the rank of Γ is bounded by $d \leq K$ and the number of nondegenerate solutions of (2.2) in Γ is bounded by $e^{c_m K}$. We now obtain the subspace theorem under the product set assumption.

Notation. $d <_h f$ means $d \leq c(h) f$, where $c(h)$ is a function of h.

Theorem 2.3 ([C2])**.** *Let $A \subset \mathbb{C}$ with $|A| = N$, and*

$$|A^2| < K|A|. \tag{2.5}$$

Then

$$|\{\textit{solutions of } x_1 + \cdots + x_{2h} = 0 \textit{ in } A\}| <_h N^{h-1}e^{cK} + N^h.$$

Theorem 2.3 gives N^3 as a bound on the number of solutions in A with $|A| = N$ to the equation

$$\xi_1 + \xi_2 + \xi_3 = \xi_4 + \xi_5 + \xi_6. \tag{2.6}$$

On the other hand, we expect (2.1) to be bounded by N^2. So we introduce a new variable z in (2.1), and let

$$x' = u'/z, \quad x = u/z,$$

where $u, u' \in X^2$. Then the equation in (2.1) becomes

$$u'y + bu' + ayz = uy' + bu + ay'z. \tag{2.7}$$

A solution $(\xi_1, \ldots, \xi_6) \in X^2Y \times bX^2 \times aXY \times X^2Y \times bX^2 \times aXY$ of (2.6) is in one-to-one correspondence to a solution $(u', u, y', y, z) \in X^2 \times X^2 \times Y \times Y \times X$ of (2.7) by the following relations:

$$\xi_1 = u'y, \quad \xi_2 = bu', \quad \xi_3 = ayz, \quad \xi_4 = uy', \quad \xi_5 = bu, \quad \xi_6 = ay'z,$$

or

$$u' = \frac{\xi_2}{b}, \quad u = \frac{\xi_5}{b}, \quad y' = \frac{b\xi_4}{\xi_5}, \quad y = \frac{b\xi_1}{\xi_2}, \quad z = \frac{\xi_2\xi_3}{ab\xi_1}.$$

In order to apply Theorem 2.3, we take

$$A = X^2Y \cup bX^2 \cup aXY.$$

Then we have $|A^2| < K|A|$ by the following Proposition 2.26 in [TV].

Proposition. *Let A, B be subsets of an abelian group with $|A| = |B| = N$. If $|A + B| < cN$, then*

$$|n_1A - n_2A + n_3B - n_4B| < c'N.$$

3. The proof of Theorem 1 for finite points

If we replace assumption (0.3) by assumption (0.1), then instead of (1.4) and Theorem 2.1, we have (3.1) and Theorem 3.1 below

$$n^{(1-\delta)/2} < |C| = |D| < n^{(1+\delta)/2}, \quad |C^{-1} \underset{\mathcal{G}}{\times} D| < n^{(1+\delta)/2}. \tag{3.1}$$

Theorem 3.1. *Let $X = \{x_i \in \mathbb{C}^2 : 1 \le i \le N^2\}$ and $Y = \{y_i \in \mathbb{C}^2 : 1 \le i \le N^2\}$ with*

$$N^{1-\delta} < |X| = |Y| < N^{1+\delta} \tag{3.2}$$

and

$$\left|\frac{Y}{X}\right| < N^{1+\delta}. \tag{3.3}$$

Fix $a, b \in \mathbb{C}$. Define

$$Z = \left\{\frac{y_i + b}{x_i + a} : 1 \le i \le N^2\right\}.$$

Then $|Z| > N^{1+\eta}$ for some $\eta = \eta(\delta) > \delta$.

Remark 3.2. Let δ' be the δ in (3.1). Then the δ in Theorem 3.1 is $(2c + 1)\delta'$ with an absolute constant c as in Theorem BSG-LR.

Similar to the argument from (2.1) to (2.7), we need to prove

$$E := |\{(u, u', y, y', z) \in X^2 \times X^2 \times Y \times Y \times X : u'y + bu' + ayz = uy' + bu + ay'z\}| < N^{4-\eta} \tag{3.4}$$

for some $\eta > 0$.

Rewriting the equation in (3.4) as

$$(y + b)u' - (y' + b)u + a(y - y')z = 0, \tag{3.5}$$

we see that (u', u) lies on the line $\ell_{y,y',z}$ defined by

$$S - \frac{y'+b}{y+b}T + \frac{a(y-y')z}{y+b} = 0. \tag{3.6}$$

Assume

$$E > N^{4-\eta}. \tag{3.7}$$

We will get a contradiction for η small. (See (3.14).)

We define

$$K = \{(y, y', z) \in Y \times Y \times X : |\ell_{y,y',z} \cap (X^2 \times X^2)| > N^{1-2\eta}\}. \tag{3.8}$$

Claim 1. *If* $3\delta < \eta$, *then*

$$|K| > \frac{E}{|X^2|}\ . \tag{3.9}$$

Proof. By (3.4)–(3.6) and (3.8),

$$E \le \sum_{y',y,z} |\ell_{y,y',z} \cap (X^2 \times X^2)| < |X^2|\,|K| + N^{1-2\eta}|X|\,|Y|^2,$$

and by (3.2), $N^{1-2\eta}|X|\,|Y|^2 < N^{1-2\eta+3(1+\delta)} < N^{4-\eta}$. The claim follows from (3.7).

Ruzsa's Inequality ([R2]). *Let M and N be finite subsets of an abelian group such that*

$$|M + N| \le \rho|M|.$$

Let $h \ge 1$ and $\ell \ge 1$. Then

$$|hN - \ell N| \le \rho^{h+\ell}|M|.$$

It follows from Ruzsa's inequality, (3.2) and (3.3) that

$$|X^2| < \left(\frac{N^{1+\delta}}{|X|}\right)^3 |X| < \frac{N^{3+3\delta}}{N^{2-2\delta}} = N^{1+5\delta}. \tag{3.10}$$

By (3.9), (3.7) and (3.10), we have

$$|K| > \frac{N^{4-\eta}}{N^{1+5\delta}} = N^{3-\eta-5\delta}. \tag{3.11}$$

Let

$$\mathcal{L} = \{\ell_{y,y',z} : (y, y', z) \in K\}. \tag{3.12}$$

Since for any (ξ, ς), there are at most $|Y| < N^{1+\delta}$ triples (y, y', z) such that

$$\xi = \frac{y'+b}{y+b}, \qquad \varsigma = \frac{a(y-y')z}{y+b},$$

for each line in $\mathcal{L}$ there are at most $N^{1+\delta}$ triples in K corresponding to it. Therefore,

$$|\mathcal{L}| > N^{2-\eta-6\delta}. \tag{3.13}$$

The following version of the Szemerédi–Trotter theorem over $\mathbb{C}$ is exactly what we need.

Szemerédi–Trotter Theorem ([S]). *Let $\mathcal{P} = C \times D \subset \mathbb{C}^2$ be a set of points and $\mathcal{L}$ be a set of lines such that $|\ell \cap \mathcal{P}| \geq k$ for any $\ell \in \mathcal{L}$. Then*

$$|\mathcal{P}|^2 > ck^3|\mathcal{L}|.$$

In the above theorem we take $\mathcal{P} = X^2 \times X^2$, $\mathcal{L}$ as in (3.12) and $k = N^{1-2\eta}$. Together with (3.10) and (3.13), we have

$$N^{4(1+5\delta)} > |X^2|^4 > c(N^{1-2\eta})^3|\mathcal{L}| > N^{5-7\eta-6\delta}.$$

This cannot happen if

$$\eta < \frac{1-26\delta}{7}. \tag{3.14}$$

Remark 3.3. The conditions that $\eta > 3\delta$ (cf. Claim 1) and (3.14) imply $\delta < 1/47$.

Remark 3.4. The case of $P_i, Q_j \in \mathbb{F}_p \times \mathbb{F}_p$ can be taken care of by the following theorem (see [B, Theorem 2.2]).

Szemerédi–Trotter Theorem for $\mathbb{F}_p$. *Let $\mathcal{P} \subset \mathbb{F}_p$ be a set of points, and $\mathcal{L}$ be a set of lines such that*

$$|\mathcal{P}|, |\mathcal{L}| \leq M < p^\alpha \quad \text{for some } 0 < \alpha < 2. \tag{3.15}$$

Let $\mathcal{I} = \{(p, \ell) \in \mathcal{P} \times \mathcal{L} : p \in \ell\}$ be the incidence relation. Then

$$|\mathcal{I}| < cM^{3/2-\gamma} \quad \text{for some } \gamma = \gamma(\alpha) > 0. \tag{3.16}$$

In (3.15), take $\mathcal{P} = X^2 \times X^2$, $\mathcal{L}$ as in (3.12), and $M = N^{2+10\delta}$ (cf. (3.10)). By (3.13) (which follows from the assumption that $E > N^{4-\eta}$), we may assume $|\mathcal{L}| = N^{2-\eta-6\delta}$. Since each line in $\mathcal{L}$ contains at least $N^{1-2\eta}$ points, we have

$$|\mathcal{I}| \geq |\mathcal{L}|N^{1-2\eta}. \tag{3.17}$$

Hence

$$cN^{(2+10\delta)(3/2-\gamma)} > N^{2-\eta-6\delta}N^{1-2\eta}.$$

This is a contradiction if δ and η are small. Therefore (3.4) holds, and Theorem 3.1 is true over $\mathbb{F}_p$.

Remark 3.5. The finite points case of Theorem 1 over $\mathbb{R}$ also follows from the following theorem by Elekes, Nathanson and Ruzsa [ENR].

Theorem ENR. *Let $S \subset \mathbb{R}$ be finite and let f be a piecewise convex function (i.e. $f' > 0$). Then*

$$|2S| + |2f(S)| \geq c|S|^{5/4}.$$

Proof of Remark 3.5. Similar to the way we derive the assumption of Theorem 3.1, we will start with (3.1) and use Theorem BSG-LR (twice, this time). Let

$$\mathcal{G} = \{(c_i, d_i) \in C \times D : 1 \le i \le N^2\}. \tag{3.18}$$

Assume

$$N^{1-\delta} < |C| = |D| < N^{1+\delta}, \quad |\mathcal{G}| \sim N^2, \tag{3.19}$$

$$\left|\left\{\frac{d_i}{c_i} : (c_i, d_i) \in \mathcal{G}\right\}\right| < N^{1+\delta}, \tag{3.20}$$

$$\left|\left\{\frac{d_i + b}{c_i + a} : (c_i, d_i) \in \mathcal{G}\right\}\right| < N^{1+\eta}. \tag{3.21}$$

First, from (3.20), we obtain $C' \subset C$ and $D' \subset D$ such that

$$|C'| \sim |C|, \quad |D'| \sim |D|, \quad |\mathcal{G} \cap (C' \times D')| \sim N^2$$

and

$$\left|\frac{D'}{C'}\right| \lesssim N^{1+\delta}. \tag{3.22}$$

Let

$$\mathcal{G}' = \mathcal{G} \cap (C' \times D').$$

Applying Theorem BSG-LR again, we obtain $X \subset C' \subset C$ and $Y \subset D' \subset D$ such that

$$|X| \sim |C'| \sim |C|, \quad |Y| \sim |D'| \sim |D|, \quad |\mathcal{G}' \cap (X \times Y)| \sim N^2,$$

$$\left|\frac{Y}{X}\right| \le \left|\frac{D'}{C'}\right| \lesssim N^{1+\delta}, \tag{3.23}$$

$$\left|\frac{Y + b}{X + a}\right| \lesssim N^{1+\eta}. \tag{3.24}$$

The bound (3.23) implies that

$$|\log Y - \log X| \lesssim N^{1+\delta}. \tag{3.25}$$

Ruzsa's inequality and (3.25) give

$$|2 \log X| \lesssim N^{1+5\delta}. \tag{3.26}$$

Assume $\delta < 1/20$. In Theorem ENR, we take $S = \log X$, and let f be the convex function $f(s) = \log(e^s + a)$. Then

$$|2 \log(X + a)| > N^{5/4}. \tag{3.27}$$

On the other hand, (3.24) implies

$$|\log(Y+b)-\log(X+a)| \lesssim N^{1+\eta}. \tag{3.28}$$

Again, applying Ruzsa's inequality to (3.28) gives

$$|2\log(X+a)| \lesssim N^{1+5\eta},$$

which contradicts (3.27) if $\eta < 1/20$.

4. The cases of points at infinity

In this section we handle all the cases when more than two of the P_i's are at infinity.

Let $P = (1, -1/d, 0) \in L_\infty$. Then the lines $L(P, Q_i)$ are defined by

$$x + dy - (c_i + dd_i)z = 0.$$

To prove Theorems 1 and 2, we need the following two theorems.

Theorem 4.1. *Let $X = \{x_i \in \mathbb{C}^2 : 1 \le i \le N^2\}$ and $Y = \{y_i \in \mathbb{C}^2 : 1 \le i \le N^2\}$ with*

$$N^{1-\delta} < |X| = |Y| < N^{1+\delta} \tag{4.1}$$

and

$$\left|\frac{Y}{X}\right| < N^{1+\delta}. \tag{4.2}$$

Fix $d \in \mathbb{C}$. Define

$$Z = \{x_i + dy_i : 1 \le i \le N^2\}. \tag{4.3}$$

Then

$$|Z| > N^{1+\eta} \quad \text{for some } \eta = \eta(\delta) \ge \delta. \tag{4.4}$$

Theorem 4.2. *Let $X = \{x_i \in \mathbb{C}^2 : 1 \le i \le N^2\}$ and $Y = \{y_i \in \mathbb{C}^2 : 1 \le i \le N^2\}$ with*

$$|X| = |Y| = c'N \quad \text{and} \quad \left|\frac{Y}{X}\right| < cN.$$

Fix $d \in \mathbb{C}$. Define $Z = \{x_i + dy_i : 1 \le i \le N^2\}$. Then $|Z| > \delta N^2$ for some $\delta > 0$.

To prove Theorem 4.1, we assume the contrary that

$$|Z| < N^{1+\eta} \tag{4.5}$$

for some $\eta = \eta(\delta) \ge \delta$. We will show that this cannot happen if η is small.

Let $A = X$, $B = dY$, where X, Y satisfy the assumptions of Theorem 4.1. Applying Theorem BSG-LR to A and B, we have

$$N^{1-\eta} < |A| = |B| < N^{1+\eta}, \tag{4.6}$$

$$\left|\frac{B}{A}\right| < N^{1+\eta}, \tag{4.7}$$

$$|A+B| < N^{1+\eta}. \tag{4.8}$$

By the same argument as that to obtain (3.10), (4.6)–(4.8) implies

$$|2A|, |A^2| < N^{1+5\eta}.$$

On the other hand, (4.6) and the sum-product theorem below imply

$$|2A| + |A^2| > N^{\frac{14}{11}(1-\eta)}.$$

This is a contradiction if $\eta < 1/23$.

Theorem (Solymosi [S])**.**

$$|2A| + |A^2| > |A|^{\frac{14}{11}-\epsilon}.$$

Remark 4.3. Let η' be the η in (4.5). Then the η in (4.6)–(4.8) is bounded by $c\eta'$, where $c \le 8$ is an absolute constant. (See Remark 1.2.) For example, if $\eta' = \delta$, we can take $\eta \le (2c+1)\delta$.

The proof of Theorem 4.2 by using the subspace theorem is rather straightforward, since as in the proof of Theorem 2.1, it suffices to show that

$$|\{(x, x', y, y') \in X \times X \times Y \times Y : x + dy = x' + dy'\}| < \frac{1}{\delta}N^2.$$

Proof of Theorem 3. Since P_1, P_2, P_3 are collinear, we may assume that $P_1 = (1, 0, 0)$, $P_2 = (0, 1, 0)$, $P_3 = (1, -1, 0) \in L_\infty$. Assumption (0.5) means that $|C|, |D|, |C + D| \lesssim N$. For a point $P = (-a, -b, 1) \notin L_\infty$, the family of lines $\{L(P, Q_j)\}_j$ corresponds to $\{\frac{d_i+b}{c_i+a} : (c_i, d_i) \in C \times D,\ 1 \le i \le N^2\}$. Applying the theorems below to the sets $C + a$, $D + b$, and by Ruzsa's inequality, we have $|(C + a)(D + b)| \sim N^{2-\epsilon}$ (respectively, $N^2/\log N$). This together with the Balog–Szemerédi–Gowers theorem implies that $|\{L(P, Q_j)\}_j| \gtrsim N^{2-\epsilon}$ (respectively, $N^2/\log N$).

Theorem ([C1])**.** *Let $A \subset \mathbb{C}$ be a finite set with $|2A| \sim |A|$. Then*

$$|A^2| > |A|^{2-\epsilon} \quad \textit{for some } \epsilon > 0.$$

Theorem (Elekes–Ruzsa [ER])**.** *Let $A \subset \mathbb{R}$ be a finite set. Then*

$$|A + A|^4 \cdot |A^2| \cdot \log|A| > |A|^6.$$

The special case of Theorem 1. Assume (0.2) holds. Then $P_1, \dots, P_4$ are collinear. After a Möbius transformation, we may assume that the four points are $P_1 = (1, 0, 0)$, $P_2 = (1, -1, 0)$, $P_3 = (0, 1, 0)$, $P_4 = (1, -1/d, 0) \in L_\infty$. The lines $\{L(P_i, Q_j)\}_j$ for $i = 1, \dots, 4$ correspond to C, $C + D$, D and $\{c_i + dd_i : (c_i, d_i) \in C \times D, 1 \le i \le N^2\}$ respectively. Since $|C| \sim |D| \sim |C + D| \sim N$, we have $C' \subset C$ with $|C'| \sim N$ and $C' \subset a + D$ for some a. Hence $C' + dD \subset a + (D + dD)$ and our conclusion follows from the following theorem.

Theorem (Konyagin–Laba [KL])**.** *Let $t \in \mathbb{C}$ be transcendental. Then*

$$|A + tA| > \frac{|A| \log |A|}{\log \log |A|}.$$

5. Higher dimensional cases

The case of $\mathbb{C}^k$ with $k > 2$ follows easily from the case of $k = 2$.

Theorem 5.1. *There is $\delta > 0$ such that for any $P_1, \dots, P_{k+2}$, $Q_1, \dots, Q_n \in \mathbb{C}^k$, if*

$$|\{L(P_i, Q_j) : 1 \le i \le k+2,\ 1 \le j \le n\}| \le n^{(k-1+\delta)/k}, \tag{5.1}$$

then $P_1, \dots, P_{k+2}$ lie on a hyperplane.

Theorem 5.2. *Given $c > 0$, there is $\delta > 0$ such that for any $P_1, \dots, P_{k+1} \in \mathbb{C}^k$ not contained in any hyperplane, and any $Q_1, \dots, Q_n \in \mathbb{C}^k$, if*

$$|\{L(P_i, Q_j) : 1 \le i \le k+1,\ 1 \le j \le n\}| \le cn^{(k-1)/k}, \tag{5.2}$$

then for any $P \in \mathbb{C}^k \setminus \{P_1, \dots, P_{k+1}\}$ we have

$$|\{L(P, Q_j) : 1 \le j \le n\}| = \delta n. \tag{5.3}$$

The set-up is similar to that of the $\mathbb{C}^2$ case. We work on $\mathbb{CP}^k$ instead of $\mathbb{C}^k$. Assuming $P_1, \dots, P_{k+1}$ are not contained in any hyperplane, after a linear transformation we may assume that $P_1 = (1, 0, \dots, 0)$, $P_2 = (0, 1, 0, \dots, , 0), \dots, P_{k+1} = (0, \dots, 0, 1)$. By the same reasoning as before, we may assume that the Q_j's all lie in the affine space. Hence we may set

$$Q_j = (c_1, \dots, c_k)^{(j)} := (c_1^{(j)}, \dots, c_k^{(j)}) \in \mathbb{R}^k \subset \mathbb{C}^k,$$

where $j = 1, \dots, n$.

Let $N = n^{1/k}$. Assumption (5.2) implies

$$|\{(c_2, \dots, c_k)^{(j)}\}_{j=1}^{N^k}|, |\{(c_1, c_3, \dots, c_k)^{(j)}\}_{j=1}^{N^k}|, \dots, |\{(c_1, \dots, c_{k-1})^{(j)}\}_{j=1}^{N^k}| < N^{k-1} \tag{5.4}$$

and

$$|\{(c_2/c_1, \dots, c_k/c_1)^{(j)}\}_{j=1}^{N^k}| < N^{k-1}. \tag{5.5}$$

For a finite point $P = (-a_1, \dots, -a_k, 1)$, the family of lines $\{L(P, Q_j) : 1 \le j \le N^k\}$ corresponds one-to-one to

$$Z = \left\{ \left(\frac{c_2 + a_2}{c_1 + a_1}, \dots, \frac{c_k + a_k}{c_1 + a_1} \right)^{(j)} : 1 \le j \le N^k \right\}.$$

Hence (5.3) is equivalent to

$$|Z| = \delta N^k \tag{5.6}$$

for some $\delta > 0$. Let $C_i = \{c_i^{(j)} : j = 1, \dots, N^k\}$. We will show that

$$|C_i| = cN \quad \text{for } i = 1, \dots, k. \tag{5.7}$$

For simpler notations and without losing generality, we give an argument for the case $k = 4$. Let

$$A = \{Q_1, \dots, Q_{N^4}\},$$

and let $p_{j_1 \cdots j_m}(x_1, \dots, x_4) = (x_{j_1}, \dots, x_{j_m})$ be the projection to the j_1-th, $\dots$, j_m-th coordinates.

First, we may assume

$$|p_{123}^{-1}(c_1, c_2, c_3) \cap A| \gtrsim N \quad \text{for all } (c_1, c_2, c_3) \in p_{123}(A). \tag{5.8}$$

In fact, let $A^c = \{(c_1, \dots, c_4) \in A : |p_{123}^{-1}(c_1, c_2, c_3) \cap A| = o(N)\}$. Then

$$|A^c| \le o(N)N^3 = o(N^4), \tag{5.9}$$

and A^c can be ignored.

Next, we see that for the set A considered in (5.8), the bound $|p_{124}(A)| \lesssim N^3$ implies

$$|p_{12}(A)| \lesssim N^2. \tag{5.10}$$

Indeed,

$$N^3 \gtrsim |p_{124}(A)| > |p_{12}(A)| \cdot \min_{(c_1,c_2)\in p_{12}(A)} |p_{124}(p_{12}^{-1}(c_1, c_2) \cap A)| \gtrsim |p_{12}(A)|\, N. \tag{5.11}$$

The last inequality is because of (5.8). Similarly, we have $|p_{13}(A)|, |p_{23}(A)| \lesssim N^2$.

Using (5.10) instead of (5.4), by the same reasoning as for (5.8), shrinking the set A in (5.8) a bit, we may assume

$$|p_{12}^{-1}(c_1, c_2) \cap A| \gtrsim N^2 \quad \text{for all } (c_1, c_2) \in p_{12}(A). \tag{5.12}$$

Therefore, (5.4) and (5.12) imply

$$N^3 \gtrsim |p_{134}(A)| \gtrsim |p_1(A)| \cdot \min_{c_1 \in p_1(A)} |p_{134}(p_1^{-1}(c_1) \cap A)| > |p_1(A)|\, N^2, \tag{5.13}$$

which implies

$$|C_1| = |p_1(A)| \lesssim N. \tag{5.14}$$

Similarly, we have $|C_2|, |C_3| \lesssim N$ for $|A| \sim N^4$.

Repeating this process on the set A obtained in (5.12) with different projections, we have $|C_4| = |p_4(A)| \lesssim N$. Now (5.7) follows from $N^4 \le |C_1|\,|C_2|\,|C_3|\,|C_4| \lesssim N^4$.

Getting back to the case of any $k > 2$, we let $B = \{Q_1, \dots, Q_{N^k}\}$. We will show that

$$|\{(c_i/c_1)^{(j)} : 1 \le j \le N^k\}| \sim N \quad \text{for all } i. \tag{5.15}$$

Let

$$C_{1i} = \{(c_1, c_i) \in C_1 \times C_i : |p_{1i}^{-1}(c_1, c_i) \cap B| \gtrsim N^{k-2}\}. \tag{5.16}$$

Since $|B| \sim N^k$, by the same reasoning as for (5.8) we have

$$|C_{1i}| \sim N^2. \tag{5.17}$$

Let π_i be the projection

$$\{(c_2/c_1, \dots, c_k/c_1)^{(j)} : (c_1, c_i)^{(j)} \in C_{1i}\} \to \{(c_i/c_1)^{(j)} : (c_1, c_i)^{(j)} \in C_{1i}\}.$$

The fiber of π_i at (c_1, c_2) corresponds one-to-one to $p_{1i}^{-1}(c_1, c_i) \cap B$. Hence the image of π_i has size $\lesssim N$ by (5.5). We replace B by $p_{1i}^{-1}(C_{1i}) \cap B$. (Note that (5.16) and (5.17) imply $|p_{1i}^{-1}(C_{1i}) \cap B| \sim N^k$.). We do this for each i (and shrink B a little if necessary.). Thus (5.15) is proved.

To prove (5.6), we want to show that under condition (5.15),

$$\left|\left\{(c_1, \dots, c_k, c_1', \dots, c_k') \in C_1 \times \cdots \times C_k \times C_1 \times \cdots \times C_k : \frac{c_i + a_i}{c_1 + a_1} = \frac{c_i' + a_i}{c_1' + a_1}, \forall i\right\}\right| \lesssim N^k. \tag{5.18}$$

It follows from the case of $\mathbb{C}^2$ that

$$\frac{c_2 + a_2}{c_1 + a_1} = \frac{c_2' + a_2}{c_1' + a_1} \tag{5.19}$$

has $\lesssim N^2$ solutions in c_1, c_2, c_1', c_2'. Fixing c_1, c_1', the equation

$$\frac{c_3 + a_3}{c_1 + a_1} = \frac{c_3' + a_3}{c_1' + a_1} \tag{5.20}$$

has at most N choices of c_3 (then c_3' is determined). Hence (5.19) and (5.20) together have $\lesssim N^3$ solutions in $c_1, c_2, c_3, c_1', c_2', c_3'$. Therefore, (5.18) follows by induction, and the finite point case of Theorem 5.2 is proved.

Only set theory is used in the argument above, hence Theorem 5.1, the other case of Theorem 5.2, and the case of $\mathbb{F}_p$ are proved in exactly the same way.

Remark 5.3. Theorems 5.1 and 5.2 are true if we replace $\mathbb{C}^k$ by $\mathbb{F}_p^k$.

6. Theorem 2 over $\mathbb{Q}$

We have a stronger result by using the λ_q constant, when the points are in $\mathbb{Q}^2$.

Theorem 6.1. *Given $\epsilon > 0$, there is $\delta > 0$ such that for any $P_1, P_2, P_3 \in \mathbb{Q}^2$ non-collinear, and $Q_1, \dots, Q_n \in \mathbb{Q}^2$, if*

$$|\{L(P_i, Q_j) : 1 \le i \le 3,\ 1 \le j \le n\}| \le n^{1/2+\epsilon}, \tag{6.1}$$

then for any $P \in \mathbb{Q}^2 \setminus \{P_1, P_2, P_3\}$, we have

$$|\{L(P, Q_j) : 1 \le j \le n\}| > n^{1-\delta}. \tag{6.2}$$

We use the same set-up as for the $\mathbb{C}$ case. Given a set $A \subset \mathbb{Q}$ with $N^{1-\epsilon} < |A| < N^{1+\epsilon}$ and $|A^2| < N^{1+5\epsilon}$, we want to bound the number of solutions $\xi_1, \ldots, \xi_6 \in A$ in the following equation by $N^{3+\delta}$ for some $\delta(\epsilon) > 0$:

$$\xi_1 + \xi_2 + \xi_3 = \xi_4 + \xi_5 + \xi_6. \tag{6.3}$$

We use the λ_q constant of A for this. We recall

Definition. *Let $A \subset \mathbb{Z}$ be finite. The* λ_q constant *of A is*

$$\lambda_{q,A} = \frac{\| \sum_{a \in A} e(ax) \|_q}{\sqrt{|A|}}, \quad \textit{where} \quad e(\theta) = e^{2\pi i \theta}.$$

Proposition ([BC])**.** *Given $\varepsilon > 0$ and $q > 2$, there exists $\delta = \delta(q, \varepsilon)$ such that if $A \subset \mathbb{Z}$ with $|A^2| < |A|^{1+\varepsilon}$, then*

$$\lambda_q(A) < |A|^{\delta},$$

where $\delta \to 0$ as $\varepsilon \to 0$. Therefore, $\| \sum_{a \in A} e(ax) \|_q < |A|^{1/2+\delta_6}$.

Define $r(\eta) = |\{(\xi_1, \xi_2, \xi_3) \in A \times A \times A : \eta = \xi_1 + \xi_2 + \xi_3\}|$. In the proposition above, we take $q = 6$. Then

$$\begin{aligned} |\{(\xi_1, \ldots, \xi_6) : \xi_1 + \xi_2 + \xi_3 = \xi_4 + \xi_5 + \xi_6\}| &= \sum r(\eta)^2 \\ &= \Big\| \Big(\sum_{a \in A} e(ax) \Big)^3 \Big\|_2^2 = \Big\| \sum_{a \in A} e(ax) \Big\|_6^6 < (N^{(1+\epsilon)(1/2+\delta_6)})^6 = N^{3+\delta}. \end{aligned}$$

References

[Bi] Bilu, Y.: Structure of sets with small sumset. In: Structure Theory of Set Addition, Astérisque **258**, 77–108 (1999) Zbl 0946.11004 MR 1701189

[B] Bourgain, J.: More on the sum-product phenomenon in prime fields and its applications. Internat. J. Number Theory **1**, 1–32 (2005) Zbl pre02205594 MR 2172328

[BC] Bourgain, J., Chang, M.-C.: On the size of k-fold sum and product sets of integers. J. Amer. Math. Soc. **17**, 473–497 (2004) Zbl 1034.05003 MR 2051619

[C1] Chang, M.-C.: Factorization in generalized arithmetic progressions and applications to the Erdős–Szemerédi sum-product problems. Geom. Funct. Anal. **13**, 720–736 (2003) Zbl 1029.11006 MR 2006555

[C2] Chang, M.-C.: Sum and product of different sets. Contrib. Discrete Math. **1**, 47–56 (2006) Zbl pre05043460 MR 2212138

[ENR] Elekes, G., Nathanson, M. B., Ruzsa, I. Z.: Convexity and sumsets. J. Number Theory **83**, 194–201 (1999)

[ER] Elekes, G., Ruzsa, I. Z.: Few sums, many products. Studia Sci. Math. Hungar. **40**, 301–308 (2003) Zbl pre05078226 MR 2036961

[ESS] Evertse, J.-H., Schlickewei, H., Schmidt, W.: Linear equations in variables which lie in a multiplicative group. Ann. of Math. **155**, 807–836 (2002)

[Fr] Freĭman, G.: Foundations of a Structural Theory of Set Addition. Transl. Math. Monogr. 37, Amer. Math. Soc. (1973) Zbl 0271.10044 MR 0360496

[KL] Konyagin, S., Laba, I.: Distance sets of well-distributed planar sets for polygonal norms. Israel J. Math. **152**, 157–179 (2006) MR 2214458

[LR] Laczkovich, M., Ruzsa, I.: The number of homothetic subsets. In: The Mathematics of Paul Erdős, II, R. L. Graham and J. Nešetril (eds.), Algorithms Combin. 14, Springer, 294–302 (1997) Zbl 0871.52012 MR 1425222

[R1] Ruzsa, I. Z.: Generalized arithmetical progressions and sumsets. Acta Math. Hungar. **65**, 379–388 (1994) Zbl 0816.11008 MR 1281447

[R2] Ruzsa, I. Z.: Sums of finite sets. In: Number Theory: New York Seminar, D. V. Chudnovsky et al. (eds.), Springer, 281–293 (1996) Zbl 0869.11011 MR 1420216

[S] Solymosi, J.: On the number of sums and products. Bull. London Math. Soc. **37**, 491–494 (2005) Zbl 1092.11018 MR 2143727

[SSV] Sudakov, B., Szemerédi, E., Vu, V.: On a question of Erdős and Moser. Duke Math. J. **129**, 129–155 (2005) Zbl pre02207898 MR 2155059

[TV] Tao, T., Vu, V.: Additive Combinatorics. Cambridge Univ. Press (2006) Zbl pre05066399

Discrete Comput Geom (2010) 43: 393–401
DOI 10.1007/s00454-009-9179-x

dc_52_10

On a Question of Erdős and Ulam

Jozsef Solymosi · Frank de Zeeuw

Received: 18 June 2008 / Revised: 10 April 2009 / Accepted: 11 April 2009 /
Published online: 7 May 2009

Abstract Ulam asked in 1945 if there is an everywhere dense *rational set*, i.e., 1 a point set in the plane with all its pairwise distances rational. Erdős conjectured that if a set S has a dense rational subset, then S should be very special. The only known types of examples of sets with dense (or even just infinite) rational subsets are lines and circles. In this paper we prove Erdős' conjecture for algebraic curves by showing that no irreducible algebraic curve other than a line or a circle contains an infinite rational set.

Keywords Rational distances · Erdős problems in discrete geometry · Rational points

1 Introduction

We define a *rational set* to be a set $S \subset \mathbb{R}^2$ such that the distance between any two elements is a rational number. We are interested in the existence of infinite rational distance sets on algebraic curves.

On any line, one can easily find an infinite rational set that is in fact dense. It is also an easy exercise to find an everywhere dense rational subset of the unit circle. On the other hand, it is not known if there is a rational set with 8 points *in general position*, i.e., no 3 on a line, no 4 on a circle. In 1945, Anning and Erdős [1] proved that any infinite *integral* set, i.e., where all distances are integers, must be contained in a line. Problems related to rational and integral sets became one of Erdős' favorite subjects in combinatorial geometry [6–9, 11, 12].

The first author was supported by NSERC and OTKA grants and by Sloan Research Fellowship.

J. Solymosi (✉) · F. de Zeeuw
Department of Mathematics, UBC, 1984 Mathematics Road, Vancouver, BC V6T 1Z2, Canada
e-mail: solymosi@math.ubc.ca

F. de Zeeuw
e-mail: fdezeeuw@math.ubc.ca

In 1945, when Ulam heard Erdős' simple proof [5] of his theorem with Anning, he said that he believed there is no everywhere dense rational set in the plane, see Problem III.5 in [22] and also [10]. Erdős conjectured that an infinite rational set must be very restricted, but that it was probably a very deep problem [10, 11]. Not much progress has been made on Ulam's question. There were attempts to find rational sets on parabolas [3, 4], and there were some results on integral sets, in particular bounds were found on the diameter of integral sets [15, 21]. Very recently Kreisel and Kurz [18] found an integral set with 7 points in general position.

In this paper, we prove that lines and circles are the only irreducible algebraic curves that contain infinite rational sets. Our main tool is Faltings' Theorem [13]. We will also show that if a rational set S has infinitely many points on a line or on a circle, then all but 4 resp. 3 points of S are on the line or on the circle. This answers questions of Guy, Problem D20 in [14], and Pach, Sect. 5.11 in [2].

2 Main Result

Our main result is the following.

Theorem 2.1 *Every rational set of the plane has only finitely many points in common with an algebraic curve defined over* $\mathbb{R}$*, unless the curve has a component which is a line or a circle.*

The two special cases, line and circle, are treated in more detail in the next theorem.

Theorem 2.2 *If a rational set S has infinitely many points on a line or on a circle, then all but* 4 *resp.* 3 *points of S are on the line or on the circle.*

Note that there are infinite rational sets with all but 4 points on a line, and there are infinite rational sets with all but 3 points on a circle. The circle case follows from the line case by applying an inversion with rational radius and center one of the 4 points not on the line. A construction of Huff [16, 19] gives an infinite rational set with all but 4 points on a line.

We can formulate our Theorem 2.1 in a different way by using the term *curve-general position*: we say that a point set S of $\mathbb{R}^2$ is in curve-general position if no algebraic curve of degree d contains more than $d(d+3)/2$ points of S. Note that $d(d+3)/2$ is the number of points in general position that determine a unique curve of degree d.

Corollary 2.3 *If S is an infinite rational set in general position, then there is an infinite $S' \subset S$ such that S' is in curve-general position.*

Proof Let S_5 consist of any five points in S, and let T_5 be the set of finitely many points on the unique conic through those five points. Continue recursively; at step n, add a point from $S \backslash T_{n-1}$ to S_{n-1} to get S_n. For each d such that $d(d+3)/2 \leq n$, let T_n be the union of T_{n-1} and the set of points of S that are on a curve of degree

d through any $d(d+3)/2$ points in S_n. Since each T_n is finite, we can always add another point. Then the infinite union of the sets S_n is an infinite subset of S with the required property. □

3 Proof of Theorem 2.1

3.1 General Approach

We will use the following theorem of Faltings [13].

Theorem (Faltings) *A curve of genus* ≥ 2, *defined over a number field*, *contains only finitely many rational points*.

In this paper by *curve* (defined over a field $K \subset \mathbb{R}$) we usually mean the zero set in $\mathbb{R}^2$ of a polynomial in two variables with coefficients from K. However, when we consider the genus of a curve, we are actually talking about the projective variety defined by the polynomial. For definitions, see [20].

First suppose that we have an infinite rational set S contained in a curve C of genus ≥ 2, defined over $\mathbb{R}$. We can move two points in S to $(0,0)$ and $(0,1)$, so that by Lemma 3.2 below the elements of S are of the form $(r_1, r_2\sqrt{k})$. Then by the remark after Lemma 3.2, the curve is defined over $\mathbb{Q}(\sqrt{k})$. By Faltings' theorem, S must be finite.

Below we will show that if we have an infinite rational set S on a curve C_1 of genus 0 or 1, then all but finitely many of the points in S will in fact give points on a curve C_2 in $\mathbb{R}^3$ of genus ≥ 2. More precisely, assuming that $(0,0)$ and $(0,1)$ are in S, a point $(r_1, r_2\sqrt{k})$ will give a point $(r_1, r_2\sqrt{k}, r_3)$ on a curve C_2, with all the r_i rational. Again we conclude by Faltings' theorem that the original set S could not have been infinite.

3.2 Two Lemmata

Rationality of distances in $\mathbb{R}^2$ is clearly preserved by translations, rotations, and uniform scaling ($(x,y) \mapsto (\lambda x, \lambda y)$ with $\lambda \in \mathbb{Q}$). More surprisingly, rational sets are preserved under certain central inversions. This will be an important tool in our proof below.

Lemma 3.1 *If we apply inversion to a rational set* S, *with center a point* $x \in S$ *and rational radius*, *then the image of* $S\backslash\{x\}$ *is a rational set*.

Proof We may assume the center to be the origin and the radius to be 1. The properties of inversion are most easily seen in complex notation, where the map is $z \mapsto 1/z$. Suppose that we have two points z_1, z_2 with rational distances $|z_1|$, $|z_2|$ from the origin and with $|z_2 - z_1|$ rational. Then

$$\left|\frac{1}{z_1} - \frac{1}{z_2}\right| = \left|\frac{z_2 - z_1}{z_1 z_2}\right| = \frac{|z_2 - z_1|}{|z_1||z_2|}$$

is also rational. dc_52_10 □

A priori, points in a rational set could take any form. However, after moving two of the points to two fixed rational points by translating, rotating, and scaling, the points are almost rational points. The following simple lemma is well known among researchers working with integer sets. As far as we know, it was proved first by Kemnitz [17].

Lemma 3.2 *For any rational set S, there is a square-free integer k such that if a similarity transformation T transforms two points of S into $(0,0)$ and $(1,0)$, then any point in $T(S)$ is of the form*

$$\left(r_1, r_2\sqrt{k}\right), \quad r_1, r_2 \in \mathbb{Q}.$$

Note that this implies that any curve of degree d containing at least $d(d+3)/2$ points from $T(S)$ is defined over $\mathbb{Q}(\sqrt{k})$.

3.3 Curves of Genus 1

Let $C_1 : f(x,y) = 0$ be an irreducible algebraic curve of genus $g_1 = 1$ and degree $d \geq 3$. Suppose that there is an infinite set S on C_1 with pairwise rational distances. Assume that the points $O = (0,0)$ and $(1,0)$ are on C_1 and in S and that O is not a singularity of C_1. Below we will be allowed to make any other assumptions on C_1 that we can achieve by translating, rotating, or scaling it, as long as we also satisfy the assumptions above. In particular, we can use any of the infinitely many rotations about the origin that put a different point of S on the x-axis.

We wish to show that the intersection curve C_2 of the surfaces

$$X : f(x,y) = 0,$$
$$Y : x^2 + y^2 = z^2,$$

has genus $g_2 \geq 2$.

Consider C_1 as a curve in the $z = 0$ plane, and define the map $\pi : C_2 \to C_1$ by $(x,y,z) \mapsto (x,y)$, i.e., the restriction to C_2 of the vertical projection from the cone Y to the $z = 0$ plane. The preimage of a point (x,y) consists of the two points $(x, y, \pm\sqrt{x^2+y^2})$, except when $x^2 + y^2 = 0$, which in $\mathbb{C}^2$ happens on the two lines $x + iy = 0$ and $x - iy = 0$. Then we can determine (or at least bound from below) the genus of C_2 using the Riemann–Hurwitz formula [20] applied to π,

$$2g_2 - 2 \geq \deg \pi \cdot (2g_1 - 2) + \sum_{P \in C_2} (e_P - 1).$$

This is usually stated with equality for smooth curves, but we are allowing C_1 and C_2 to have singularities. To justify our use of it, observe that the map π corresponds to a map $\tilde{\pi} : \tilde{C}_1 \to \tilde{C}_2$ between the normalizations of the curves, for which Riemann–Hurwitz holds. The normalizations have the same genera as the original curves, and

$\tilde{\pi}$ has the same degree. Furthermore a ramification point of π away from any singularities gives a ramification point of $\tilde{\pi}$. It is enough for our purposes to have this inequality, but there could be more ramification points for $\tilde{\pi}$, above where the singularities used to be.

Applying this formula with $g_1 = 1$, $d = 2$, we have

$$g_2 \geq 1 + \frac{1}{2} \sum_{P \in C_2} (e_P - 1),$$

so to get $g_2 \geq 2$, we only need to show that π has some ramification point.

The potential ramification points are above the intersection points of C_1 with the lines $x \pm iy = 0$, of which there are $2d$ by Bézout's theorem, counting with multiplicities. Such an intersection point P can only fail to have a ramification point above it if the curve has a singularity at P or if the curve is tangent to the line there. We will show that there are only finitely many lines through the origin on which one of those two things happens. Then certainly one of the infinitely many rotations of C_1 that we allowed above will give an intersection point of C_1 with $x \pm iy = 0$ that has a ramification point above it.

The intersection of a line $y = ax$ with $f(x, y) = 0$ is given by $p_a(x) = f(x, ax) = 0$, and if the point of intersection is a singularity or a point of tangency, then $p_a(x)$ has a multiple root. We can detect such multiple roots by taking the discriminant of $p_a(x)$, which will be a polynomial in a that vanishes if and only if $p_a(x)$ has a multiple root. Hence for all but finitely many values of a, the line $y = ax$ has d simple intersection points with $f(x, y) = 0$. So indeed there is an allowed rotation after which π is certain to have a ramification point.

3.4 Curves of Genus 0, $d \geq 4$

Let $C_1 : f(x, y) = 0$ be an irreducible algebraic curve of genus $g_1 = 0$, and again assume that it passes through the origin but does not have a singularity there. Then Riemann–Hurwitz with the same map π as above gives

$$g_2 \geq -1 + \frac{1}{2} \sum_{P \in C_2} (e_P - 1),$$

so to get $g_2 \geq 2$ we need to show that there are at least 5 ramification points. As above, we can ensure that the lines $x \pm iy$ each have d simple points of intersection. Discounting the intersection point of the two lines, this gives $2d - 2$ ramification points. Hence if the degree of f is $d \geq 4$, we are done.

3.5 Curves of Genus 0, $d = 2, 3$

Let $d = 3$ and assume that $f(x, y) = 0$ is not a line or a circle. Consider applying inversion with the origin as center to the curve. This is a birational transformation, so does not change the genus. Therefore, when inversion increases the degree of f to above 4, we are done.

Algebraically, inversion in the circle around the origin with radius 1 is given by

$$(x, y) \mapsto \left(\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2}\right),$$

and since this map is its own inverse, the curve $f(x, y) = 0$ is sent to the curve

$$C_3 : \left(x^2+y^2\right)^k \cdot f\left(\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2}\right) = 0,$$

where $k \le d$ is the lowest integer that makes this a polynomial. This curve is irreducible if and only if the original curve is irreducible. Since f does not have a singularity at the origin, it has a linear term $ax + by$ with a, b not both zero. After inversion this gives a highest-degree term

$$(ax+by)\left(x^2+y^2\right)^{k-1}.$$

In our situation, $d = 3$, so if $k = 3$, the curve C_3 has degree $2k - 1 = 5$, and we are done.

The only other possibility is that $k = 2$, which happens if $x^2 + y^2$ divides the leading terms of f. We will treat these cases in a completely different way.

If $d = 2$, then applying inversion will give a curve of degree 3, unless its leading terms are $x^2 + y^2$, which exactly means that it is a circle! So we treat this case by reducing it to the $d = 3$ case.

Since f has degree 3 and genus 0, it must have a singularity. The singularity need not be in our rational set, but it is always a rational point, so we can move it to the origin, while maintaining the almost-rational form of the points in our rational set. Then f must have the form

$$(ax+by)\left(x^2+y^2\right) + cx^2 + dy^2 + exy.$$

Note that this is exactly what we get if we apply inversion to a quadratic that is not a circle and goes through the origin.

In fact, we can ensure that $(1, 0)$ is on the curve again, so that $a + c = 0$. Then if we divide by c, f is of the form

$$(-x+by)\left(x^2+y^2\right) + x^2 + dy^2 + exy.$$

We can parameterize this curve using lines $x = ty$, giving the parameterization

$$y(t) = \frac{t^2+et+d}{(t-b)(t^2+1)} =: \frac{p(t)}{q(t)}, \qquad x(t) = t \cdot y(t).$$

If we let t_j be a value of t that gives one of the points from our rational distance set, it follows that for infinitely many t,

$$\left(y(t)-y(t_j)\right)^2 + \left(x(t)-y(t_j)\right)^2 = \left(\frac{p(t)}{q(t)} - \frac{p(t_j)}{q(t_j)}\right)^2 + \left(t \cdot \frac{p(t)}{q(t)} - t_j \cdot \frac{p(t_j)}{q(t_j)}\right)^2$$

is a square. Then we can multiply by $q(t)^2 q(t_j)^2$ to get infinitely many squares of the form

$$\big(p(t)q(t_j) - p(t_j)q(t)\big)^2 + \big(tp(t)q(t_j) - t_j p(t_j)q(t)\big)^2.$$

This polynomial has degree 6 in t. It has a factor $(t - t_j)^2$ and a factor $t^2 + 1$, since taking $t = \pm i$ gives (using $q(\pm i) = 0$)

$$\big(p(\pm i)q(t_j)\big)^2 + \big(\pm i \cdot p(\pm i)q(t_j)\big)^2 = 0.$$

Factoring these out, we get a quadratic polynomial $Q_j(t)$ in t. Its leading coefficient is

$$\big(t_j^2 + 1\big)\big(\big(d^2 + b^2\big)t_j^2 + 2\big(b^2 e + db - d^2 b\big)t_j + b^2 e^2 + b^2 d^2 + d^2 + 2ebd\big),$$

and its constant term is

$$\big(t_j^2 + 1\big)\big(\big(1 + (e + b)^2\big)t_j^2 + 2(bd - b + de)t_j + d^2 + b^2\big).$$

These polynomials in t_j are not identically zero (if b and d were both 0, then f would be reducible), hence we can pick t_j so that they are not zero. Then in turn $Q_j(t)$ is a proper quadratic polynomial, and since it is essentially a distance function in the real plane, it cannot have real roots, so it has two distinct imaginary roots.

Therefore our infinite rational set gives infinitely many solutions to the equations

$$z_j^2 = \big(t^2 + 1\big) \cdot Q_j(t).$$

Multiplying three of these together, and moving $(t^2 + 1)^2$ into the square on the left, we get infinitely many solutions to

$$z^2 = \big(t^2 + 1\big) Q_1(t) Q_2(t) Q_3(t).$$

If there are no multiple roots on the right, then this is a hyperelliptic curve of degree 8, so it has genus 3, hence cannot have infinitely many solutions, a contradiction.

The one thing we need to check is that we can choose the t_j so that the Q_j do not have roots in common. We need some notation: write

$$Q_j(t) = c_2(t_j)t^2 + c_1(t_j)t + c_0(t_j),$$

where

$$\begin{aligned}
c_2(t_j) &= \big(1 + (e + b)^2\big)t_j^2 + 2(bd + de - b)t_j + d^2 + b^2 \\
c_1(t_j) &= 2(bd + de - b)t_j^2 + 2\big(b^2 + d^2 - bed - bd - be - d\big)t_j \\
&\quad + 2\big(bd + b^2 e - bd^2\big) \\
c_0(t_j) &= \big(d^2 + b^2\big)t_j^2 + 2\big(b^2 e + db - d^2 b\big)t_j + b^2 e^2 + b^2 d^2 + d^2 + 2ebd.
\end{aligned}$$

Suppose that for infinitely many t_j, the polynomial $Q_j(t)$ has the same roots x_1 and x_2. Then for each of those t_j, we have

$$c_1(t_j) = -(x_1 + x_2) \cdot c_2(t_j), \qquad c_0(t_j) = x_1 \cdot x_2 \cdot c_2(t_j).$$

If we look at the coefficients of the t_j terms in these equations, we see that

$$-x_1 - x_2 = \frac{2(b^2 + d^2 - bed - bd - be - d)}{2(bd - b + de)} = -b - \frac{be + d - d^2}{bd + de - b},$$
$$x_1 \cdot x_2 = \frac{2(b^2 e + db - d^2 b)}{2(bd + de - b)} = b \cdot \frac{be + d - d^2}{bd + de - b}.$$

Here we can read off that the roots are $x_1 = b$ and $x_2 = \frac{be+d-d^2}{bd+de-b}$, which is a contradiction, since the roots had to be imaginary.

4 Proof of Theorem 2.2

We will prove that if a rational set has infinitely many points on a line, then it can have at most 4 points off the line. The corresponding statement for 3 points off a circle then follows by applying an inversion. More precisely, suppose that we have a rational set S with infinitely many points on a circle C and at least 4 points off that circle. Assume that the origin is one of the points in $S \cap C$, and apply inversion with the origin as center and with some rational radius. That turns C into a line L, and we get a rational set with infinitely many points on L and 4 other points. Moreover, the new origin can be added to S, so that we get 5 points off the line, contradicting what we will prove below. To see that the new origin has rational distance to all points in S, observe that in complex notation the distances $|z|$ to the old origin were rational for all $z \in S$ and that the distances to the new origin are $1/|z|$.

To prove the statement for a line, our main tool will again be Faltings' theorem, but now applied to the hyperelliptic curve

$$y^2 = \prod_{i=1}^{6} (x - \alpha_i),$$

which has genus 2 if and only if the α_i are distinct.

Suppose that we have a rational set S with infinitely many points on a line, say the x-axis, and 5 or more points off that line. Then we can assume that 3 of those points are above the x-axis and that one of them is at $(0, 1)$. Let the other two points be at (a_1, b_1) and (a_2, b_2). Note that we are taking 3 points on one side of the line, because we want to avoid having one point a reflection of another. If we had, say, $(a_1, b_1) = (0, -1)$, the argument below would break down.

Take a point $(x, 0)$ of S on the x-axis with $x \neq 0, a_1, a_2$. Then we have that

$$x^2 + 1, \qquad (x - a_1)^2 + b_1^2, \quad \text{and} \quad (x - a_2)^2 + b_2^2$$

are rational squares, so that we get a rational point (x, y) on the curve

$$y^2 = \bigl(x^2+1\bigr)\bigl((x-a_1)^2+b_1^2\bigr)\bigl((x-a_2)^2+b_2^2\bigr).$$

This is a curve of genus 2, since the roots on the right-hand side are distinct: they are $\pm i$ and $x = a_i \pm \sqrt{-b_i^2}$ for $i = 1, 2$, which are distinct by the assumptions on the points (a_i, b_i).

Therefore the curve has genus 2 and cannot contain infinitely many rational points, contradicting the fact that S has infinitely many points on the line.

Acknowledgements We thank Kalle Karu for the useful discussions. We are also indebted to an anonymous referee who noticed that the $d = 3$ case in Sect. 3.7 was not completely covered in the previous version of the paper.

References

1. Anning, N.H., Erdős, P.: Integral distances. Bull. Am. Math. Soc. **51**, 598–600 (1945)
2. Brass, P., Moser, W., Pach, J.: Research Problems in Discrete Geometry, 1st edn. Springer, Berlin (2005). XII, 499 p.
3. Campbell, G.: Points on $y = x^2$ at rational distance. Math. Comput. **73**, 2093–2108 (2004)
4. Choudhry, A.: Points at rational distances on a parabola. Rocky Mt. J. Math. **36**(2), 413–424 (2006)
5. Erdős, P.: Integral distances. Bull. Am. Math. Soc. **51**, 996 (1945)
6. Erdős, P.: Verchu niakoy geometritchesky zadatchy. Fiz.-Mat. Spis. Bŭlgar. Akad. Nauk **5**(38), 205–212 (1962). (On some geometric problems, in Bulgarian)
7. Erdős, P.: On some problems of elementary and combinatorial geometry. Ann. Mat. Pura Appl. (IV) **CIII**, 99–108 (1975)
8. Erdős, P.: Néhány elemi geometriai problémáról. Középisk Mat. Lapok **61**, 49–54 (1980). (On some problems in elementary geometry, in Hungarian)
9. Erdős, P.: Combinatorial problems in geometry. Math. Chron. **12**, 35–54 (1983)
10. Erdős, P.: Ulam, the man and the mathematician. J. Graph Theory **9**(4), 445–449 (1985) Also appears in Creation Math. 19 (1986), 13–16
11. Erdős, P.: Some combinatorial and metric problems in geometry. In: Colloquia Mathematica Societatis János Bolyai, vol. 48, pp. 167–177. Intuitive Geometry, Siófok (1985)
12. Erdős, P., Purdy, G.B.: Extremal problems in combinatorial geometry. In: Graham, R.L., Grötschel, M., Lovász, L. (eds.) Handbook of Combinatorics, pp. 809–875. Elsevier, Amsterdam (1995)
13. Faltings, G.: Endlichkeitssätze für abelsche Varietäten über Zahlkörpern. Invent. Math. **73**(3), 349–366 (1983). (Finiteness theorems for abelian varieties over number fields)
14. Guy, R.: Unsolved Problems in Number Theory, 3rd edn. Problem Books in Mathematics Subseries: Unsolved Problems in Intuitive Mathematics, vol. 1. Springer, Berlin (2004). XVIII, 438 p.
15. Harborth, H., Kemnitz, A., Möller, M.: An upper bound for the minimum diameter of integral point sets. Discrete Comput. Geom. **9**(4), 427–432 (1993)
16. Huff, G.B.: Diophantine problems in geometry and elliptic ternary forms. Duke Math. J. **15**, 443–453 (1948)
17. Kemnitz, A.: Punktmengen mit ganzzahligen Abständen. Habilitationsschrift, TU Braunschweig (1988)
18. Kreisel, T., Kurz, S.: There are integral heptagons, no three points on a line, no four on a circle. Discrete Comput. Geom. **39**(4), 786–790 (2008)
19. Peeples, W.D. Jr.: Elliptic curves and rational distance sets. Proc. Am. Math. Soc. **5**, 29–33 (1954)
20. Silverman, J.: The Arithmetic of Elliptic Curves. Springer, Berlin (1986)
21. Solymosi, J.: Note on integral distances. Discrete Comput. Geom. **30**(2), 337–342 (2003)
22. Ulam, S.M.: A Collection of Mathematical Problems. Interscience Tracts in Pure and Applied Mathematics, vol. 8. Interscience, New York (1960). XIII, 150 p.

COMBINATORICA dc_52_10
Bolyai Society – Springer-Verlag

COMBINATORICA **28** (1) (2008) 113–125
DOI: 10.1007/s00493-008-2099-1

NEAR OPTIMAL BOUNDS FOR THE ERDŐS DISTINCT DISTANCES PROBLEM IN HIGH DIMENSIONS

JÓZSEF SOLYMOSI, VAN H. VU

Received October 23, 2003
Revised September 5, 2004

We show that the number of distinct distances in a set of n points in $\mathbb{R}^d$ is $\Omega(n^{\frac{2}{d}-\frac{2}{d(d+2)}})$, $d \geq 3$. Erdős' conjecture is $\Omega(n^{2/d})$.

1. Introduction

One of the most famous and important problems in discrete geometry is the following question, posed by Erdős [7, 1]:

What is the minimum number of distinct distances determined by n points in $\mathbb{R}^d$?

Given a finite point set A, let $g(A)$ denote the number of distinct distances between the elements of A. Define $g_d(n) = \min_{A \subset \mathbb{R}^d, |A|=n} g(A)$. Erdős' question is to estimate $g_d(n)$. To this end, d is a constant and n is sufficiently large. The asymptotic notation is used under the assumption that $n \to \infty$.

To find an upper bound for $g_d(n)$, let us consider the following natural construction. Let A be the set of integral lattice points $(x_1, \ldots, x_d)$ where $1 \leq x_i \leq n^{1/d}$, assuming that $n^{1/d}$ is an integer. The distance between any two points in A is the square root of a positive integer less than $dn^{2/d}$. This shows that $g_d(n) = O(n^{2/d})$. Erdős and many other researchers conjecture that $g_d(n)$ is close to this upper bound.

Research on this problem has led to various new methods and concepts which are very useful for many other problems in discrete and computational

Mathematics Subject Classification (2000): 52C10

dc_52_10

geometry. The monograph by Agarwal and Pach [1] is an excellent place to read about these developments. In the last few years, an intersting link was found between the Erdős distance problem and problems in analysis. The reader who is interested in this new direction is referred to a recent survey by Iosevich [8].

Let us now give a brief account about previous lower bounds of $g_d(n)$. Erdős proved, in 1946, that $g_2(n) = \Omega(n^{1/2})$ [7]. It is easy to show, using a variant of his argument that $g_d(n) = \Omega(n^{1/d})$, for all $d \geq 1$. There is a series of improvements for the case $d = 2$, due to Moser [11], Chung [4], Chung–Szemerédi–Trotter [5], Székely [14], Solymosi–Tóth [12] and Tardos [15]. The most current bound is $g_2(n) = \Omega(n^{0.8635})$ [15]. Little has been known for $d \geq 3$. Clarkson, Edelsbrunner, Gubias, Sharir and Welzl [6] proved that $g_3(n) = \Omega(n^{1/2})$. Very recently, Aronov, Pach, Sharir and Tardos [2] proved that $g_3(n) = \Omega(n^{77/141-\epsilon})$ for any positive constant ϵ. More general, they proved that $g_d(n) = \Omega(n^{1/(d-90/77)-\epsilon})$ for any $d \geq 3$. This result gives a non-trivial improvement for small d, compared to the previous bound $n^{1/d}$. On the other hand, as d is getting large, the exponent $1/(d-90/77)-\epsilon$ converges to $1/d$, rather than to the conjectured bound $2/d$.

Our main goal in this paper is to prove that the exponent $2/d$ is essentially best possible, as it cannot be replaced by $(2-\epsilon)/d$ for any positive constant ϵ, given that d is sufficiently large. More precisely, we show that $g_d(n) = \Omega(n^{(2-\epsilon_d)/d})$, where $\epsilon_d = O(1/d)$ tends to 0 as d tends to infinity. Our bound improves the above mentioned result by Aronov et al. for every $d \geq 3$.

Theorem 1.1. *(a)* $g_3(n) = \Omega(n^{.5643})$.

(b) For any $d \geq 4$, $g_d(n) = \Omega\left(n^{\frac{2}{d} - \frac{2}{d(d+2)}}\right)$.

This theorem is a corollary of the following stronger result, which gives a recursive estimate for $g_d(n)$.

Theorem 1.2. *(a) If* $g_{d_0}(n) = \Omega(n^{\alpha_{d_0}})$, *then for all* $d \geq d_0$

$$g_d(n) = \Omega\left(n^{\frac{2d}{(d+d_0+1)(d-d_0)+2d_0/\alpha_{d_0}}}\right). \tag{1}$$

(b) If $g_{d_0}(n) = \Omega(n^{\alpha_{d_0}})$, *then for all* $d \geq d_0$, $d - d_0$ *even*

$$g_d(n) = \Omega\left(n^{\frac{2(d+1)}{(d+d_0+2)(d-d_0)+2(d_0+1)/\alpha_{d_0}}}\right). \tag{2}$$

Tardos result [15] asserts that one can set $\alpha_2 = .8635$. Applying (1) with $d_0 = 2$, $d = 3$ and $\alpha_2 = .8635$ gives $g_3(n) = \Omega(n^{.5643})$, proving part (a) of Theorem 1.1. This estimate improves the bound $\Omega(n^{77/141-\epsilon})$ by Aronov et

al. as $77/141 < .5461$. This bound on $g_3(n)$ can be further improved to $n^{.566}$ using additional arguments. The details will appear later.

Part (b) of Theorem 1.2 implies:

Corollary 1.3. *For any even d,*

$$g_d(n) = \Omega\left(n^{\frac{2(d+1)}{d^2+2d-8+6/\alpha_2}}\right). \tag{3}$$

For any odd $d \geq 3$

$$g_d(n) = \Omega\left(n^{\frac{2(d+1)}{d^2+2d-15+8/\alpha_3}}\right). \tag{4}$$

As mentioned above, we can set $\alpha_2 = .8635$ and $\alpha_3 = .5643$. With these values, the exponents in Corollary 1.3 are larger than $\frac{2}{d} - \frac{2}{d(d+2)}$ in both cases. This proves part (b) of Theorem 1.1.

We would like to point out that for those d where $d - d_0$ is an even positive integer, the bound in part (b) of Theorem 1.2 is superior to the bound in part (a). We leave the details as an exercise.

Finally, let us mention that recently several variants of Erdős distance problem have been raised by analysts. The method developed in this paper helps us to obtain new results concerning these problems. The details will appear in a future paper.

The rest of the paper is organized as follows. In the next section, we present two recursive theorems and use them to obtain Theorem 1.2. The next section, Section 3, discusses a lemma that we need in the proof of these recursive theorems. The full proofs of these theorems follow in Sections 4 and 5, respectively. The final section, Section 6, is devoted to concluding remarks.

2. Recursions

For a finite set A we denote by $t(A)$ the maximum number of distinct distances measured from a point in A. Furthermore, define

$$t_d(n) = \min_{A \subset \mathbb{R}^d, |A|=d} t(A).$$

It is clear that $t_d(n) \leq g_d(n)$. Instead of lower bounding $g_d(n)$, we are going to bound $t_d(n)$ from below. All theorems and corollaries in this section hold, with the same proofs, if we replace $t_d(n)$ by $g_d(n)$.

dc_52_10

Theorem 2.1. *Let A be a set of n points in $\mathbb{R}^d$ $(d \geq 3)$ and m be the maximum cardinality of the intersection of A with a hyperplane of co-dimension 1. Then*

$$t(A) = \Omega\Big(\max\Big\{\frac{n}{m^{(d-1)/d}}, t_{d-1}(m)\Big\}\Big). \tag{5}$$

Theorem 2.2. *Let A be a set of n points in $\mathbb{R}^d$ $(d \geq 3)$ and m be the maximum cardinality of the intersection of A with a hyperplane of co-dimension 2. Then*

$$t(A) = \Omega\left(\max\left\{\frac{n^{(d+1)/2d}}{m^{(d-1)/2d}}, t_{d-2}(m)\right\}\right). \tag{6}$$

2.1. A recursion using Theorem 2.1

In this subsection, we use Theorem 2.1 to obtain part (a) of Theorem 1.2. First, we can prove the following general result.

Corollary 2.3. *Let α be a positive constant such that $t_{d-1}(n) = \Omega(n^\alpha)$, then*

$$t_d(n) = \Omega\Big(n^{\frac{d\alpha}{d\alpha+(d-1)}}\Big). \tag{7}$$

Proof. Theorem 2.1 implies that

$$t_d(n) = \Omega\Big(\frac{n}{m^{(d-1)/d}} + t_{d-1}(m)\Big) = \Omega\Big(\frac{n}{m^{(d-1)/d}} + m^\alpha\Big). \tag{8}$$

Set $\theta = \frac{d\alpha}{d\alpha+(d-1)}$. By convexity,

$$\frac{n}{m^{(d-1)/d}} + m^\alpha \geq \Big(\frac{n}{m^{(d-1)/d}}\Big)^\theta (m^\alpha)^{1-\theta} = \Omega(n^\theta) = \Omega\Big(n^{\frac{d\alpha}{d\alpha+(d-1)}}\Big), \tag{9}$$

completing the proof. ∎

Corollary 2.3 gives rise to the following recursive estimate. Assume that for some $d_0 \geq 1$ there is a constant α_{d_0} such that $t_{d_0}(n) = \Omega(n^{\alpha_{d_0}})$. Define

$$\alpha_d = \frac{d\alpha}{d\alpha_{d-1} + (d-1)} \tag{10}$$

for $d \geq d_0 + 1$.

Corollary 2.4. *With the above assumption and notation, we have*

$$t_d(n) = \Omega(n^{\alpha_d}). \tag{11}$$

We have an exact formula for α_d, given α_{d_0}.

Fact 2.5. *For any $d \geq d_0$*

$$\alpha_d = \frac{2d}{(d+d_0+1)(d-d_0)+2d_0/\alpha_{d_0}}. \tag{12}$$

Corollary 2.4 and Fact 2.5 imply statement (a) of Theorem 1.2.

Proof. Define $\gamma_d = 1/\alpha_d$; (10) implies

$$\gamma_d = 1 + \frac{d-1}{d}\gamma_{d-1}. \tag{13}$$

Using induction, it is easy to show that for any $d \geq d_0$

$$\gamma_d = \frac{(d+d_0+1)(d-d_0)}{2d} + \frac{d_0}{d}\gamma_{d_0}, \tag{14}$$

which is equivalent to (12). ∎

2.2. A recursion using Theorem 2.2

The arguments here are very similar to the arguments in the previous subsection. As an analogue of Corollary 2.3, we have:

Corollary 2.6. *Let α be a positive constant such that $t_{d-2}(n) = \Omega(n^\alpha)$, then*

$$t_d(n) = \Omega\left(n^{\frac{(d+1)\alpha}{2d\alpha+(d-1)}}\right). \tag{15}$$

Proof. Theorem 2.2 implies that

$$t_d(n) = \Omega\left(\frac{n^{(d+1)/2d}}{m^{(d-1)/2d}} + t_{d-2}(m)\right) = \Omega\left(\frac{n^{(d+1)/2d}}{m^{(d-1)/2d}} + m^\alpha\right). \tag{16}$$

Set $\theta = \frac{2d\alpha}{2d\alpha+(d-1)}$. By convexity,

$$\begin{aligned}\frac{n^{(d+1)/2d}}{m^{(d-1)/2d}} + m^\alpha \geq \left(\frac{n^{(d+1)/2d}}{m^{(d-1)/2d}}\right)^\theta (m^\alpha)^{1-\theta} &= \Omega(n^{\theta(d+1)/2d}) \\ &= \Omega\left(n^{\frac{(d+1)\alpha}{2d\alpha+(d-1)}}\right),\end{aligned} \tag{17}$$

completing the proof. ∎

Assume that for some $d_0 \geq 1$ there is a constant α_{d_0} such that $t_{d_0}(n) = \Omega(n^{\alpha_{d_0}})$. Define

$$\alpha_d = \frac{(d+1)\alpha_{d-2}}{2d\alpha_{d-2}+(d-1)}, \tag{18}$$

for $d = d_0+2, d_0+4$, etc.

dc_52_10

Corollary 2.7. *With the above assumption and notation, we have*

$$t_d(n) = \Omega(n^{\alpha_d}). \tag{19}$$

For a fixed pair of d_0 and α_{d_0}, we can give an explicit formula for α_d.

Fact 2.8. *For any $d \geq d_0$ and $d - d_0$ even*

$$\alpha_d = \frac{2(d+1)}{(d+d_0+2)(d-d_0) + 2(d_0+1)/\alpha_{d_0}}. \tag{20}$$

Proof. Define $\gamma_d = 1/\alpha_d$; (18) implies

$$\gamma_d = \frac{2d}{d+1} + \frac{d-1}{d+1}\gamma_{d-2}. \tag{21}$$

Using induction, it is easy to show that for any $d \geq d_0$ and $d - d_0$ even

$$\gamma_d = \frac{(d+d_0+2)(d-d_0)}{2(d+1)} + \frac{d_0+1}{d+1}\gamma_{d_0}, \tag{22}$$

which is equivalent to (20). ∎

Corollary 2.7 and Fact 2.8 together imply part (b) of Theorem 1.2.

3. Partition of spaces

In this section, we present a lemma which we shall need in the proofs of Theorems 2.1 and 2.2. The development of this lemma was motivated by practical problems in geometric searching. One of the main techniques for doing a search is divide-and-conquer. In many problems, the situation looks as follows: Given a set B of hyperplanes (of co-dimension 1) in $\mathbb{R}^d$, one would like to partition $\mathbb{R}^d$ in not too many parts so that each part intersects only few hyperplanes. The following lemma, due to Chazelle and Friedman [3] was discovered along these lines. The reader who is interested in this lemma and its applications is referred to Section 6 of Matousek's monograph [10], which contains a detailed discussion about this lemma and its origin.

Definition 3.1. A hyperplane H strongly intersects a set P if $H \cap P$ is not empty and P has a point on both side of H.

Lemma 3.2. *Let B be a set of k hyperplanes in $\mathbb{R}^d$. For any $1 \leq r \leq k$, one can partition $\mathbb{R}^d$ into r sets $P_1, \ldots, P_r$ such that for each $1 \leq i \leq r$, there are only $O(k/r^{1/d})$ planes which strongly intersect P_i.*

The bound $O(k/r^{1/d})$ is best possible; the hidden constants in O depend on d but not on r. One can also guarantee that the sets P_i are generalized simplices. Strong intersection actually means intersection with the interior (see [10]), but this information is not important to our proofs. Let us now consider a little bit more complex situation when beside B we also have a set A of n points. We can require, in addition, that each part contains not too many points.

Lemma 3.3. *Let A be a set of n points and B be a set of k hyperplanes in $\mathbb{R}^d$. For any $1 \le r \le k$, one can partition $\mathbb{R}^d$ into r sets $P_1, \dots, P_r$ such that for each $1 \le i \le r$, $|P_i \cap A| \le 2n/r$ and P_i strongly intersects $O(k/r^{1/d})$ planes.*

Proof. Let us assume, without loss of generality, that r is even and $2n/r$ is an integer. Apply Lemma 3.2 with $r' = r/2$. If $|P_i \cap A| \le 2n/r$ for all $i = 1, \dots, r'$ then we are done. Otherwise, for each i where $|P_i \cap A| > 2n/r$, partition P_i into smaller parts so that all but at most one of them have exactly $2n/r$ points. The resulting finer partition has at most $r' + r/2 = r$ parts and each part satisfies the requirement of the lemma. ∎

Lemma 3.2 is not restricted to hyperplanes. It is known that this lemma still holds if we replace a family of hyperplanes by a family of surfaces satisfying certain topological conditions. In particular, the lemma holds if we replace hyperplanes by (full dimensional) spheres (see Section 6.5 of [10]). As an analogue of Lemma 3.3, we obtain the following lemma, which we shall use in the next proof.

Definition 3.4. A sphere S strongly intersects a set P if $S \cap P$ is not empty and P has a point on both side of S.

Lemma 3.5. *Let A be a set of n points and B be a set of k spheres in $\mathbb{R}^d$. For any $1 \le r \le k$, one can partition $\mathbb{R}^d$ into r sets $P_1, \dots, P_r$ such that for each $1 \le i \le r$, $|P_i \cap A| = O(n/r)$ and there are only $O(k/r^{1/d})$ spheres which strongly intersect P.*

4. Proof of Theorem 2.1

Since m is the maximum number of points of A on a hyperplane, there is a hyperplane of dimension $d-1$ containing m points of A and thus $t(A) \ge t_{d-1}(m)$. The non-trivial half of the bound is to show $t(A) = \Omega(\frac{n}{m^{(d-1)/d}})$.

Set $t = t(A)$. Since there are at most t distinct distances from v, all points in A (except v) are contained on t spheres $S_1(v), \dots, S_t(v)$ centered at v (we can add a few dummy spheres which contain no points from A). Together we

have $k = nt$ spheres. We apply Lemma 3.5 to A and the collection of these k spheres. The sets $P_1, \dots, P_r$ in the partition will be referred to as *cells*.

We call a pair (u, v), $u \in A, v \in A$, *consistent* if u and v belong to the same cell. Let M_r denote the number of consistent pairs. We are going to estimate M_r from both above and below. The statement of the theorem will follow from these estimates, under a proper choice of r.

Since $|P_i \cap A| = O(n/r)$ for all $1 \le i \le r$,

$$M_r = O\left(r\binom{n/r}{2}\right) = O\left(\frac{n^2}{r}\right). \tag{23}$$

To lower bound M_r, let us consider a point $v \in A$. If a cell P has a common point with $S_i(v)$ but does not intersect $S_i(v)$ strongly, then we say that it intersects $S_i(v)$ *weakly*. Let $s_i(v)$ be the number of cells intersecting $S_i(v)$ (either strongly or weakly).

Consider a sphere $S_i(v)$. Without loss of generality, we can assume that the cells intersecting $S_i(v)$ are $P_1, \dots, P_l$. Let $x_j = |P_j \cap S_i(v)|$, $1 \le j \le l$. The number of consistent pairs on S_i is

$$\sum_{j=1}^{l} \binom{x_j}{2} \ge \sum_{x_j \ge 1} (x_j - 1) = |A \cap S_i(v)| - s_i(v). \tag{24}$$

Summing the above estimate over all spheres $S_i(v)$ centered at v and then summing over all $v \in A$ give us

$$\sum_{v \in A} \sum_{S_i(v)} (|A \cap S_i(v)| - s_i(v))$$

consistent pairs. However, this is not yet an estimate for M_r, as a pair can be counted many times. Indeed, if the vertices of a pair are of the same distance from p points in A, then the pair is counted p times. The points which are at the same distance from the vertices of a pair lie on a hyperplane. We assume that a hyperplane contains at most m points from A, so the multiplicity of any pair is at most m. It follows that

$$M_r \ge \frac{1}{m} \sum_{v \in A} \sum_{S_i(v)} (|A \cap S_i(v)| - s_i(v)). \tag{25}$$

Now we are going to bound the right hand side of (25) from below. First of all, it is trivial that for any $v \in A$

$$\sum_{S_i(v)} |A \cap S_i(v)| = |A \setminus \{v\}| = n - 1,$$

so

$$\sum_{v\in A}\sum_{S_i(v)} |A\cap S_i(v)| = n(n-1). \tag{26}$$

To estimate $\sum_{v\in A}\sum_{S_i(v)} s_i(v)$, we split each $s_i(v)$ as the sum of two terms $s_i'(v)$ and $s_i''(v)$ which count the number of strong and weak intersections, respectively. It follows that

$$\sum_{v\in A}\sum_{i=1}^{t} s_i(v) = \sum_{v\in A}\sum_{i=1}^{t} s_i'(v) + \sum_{v\in A}\sum_{i=1}^{t} s_i''(v). \tag{27}$$

The sum $\sum_{v\in A}\sum_{i=1}^{t} s_i'(v)$ counts the total number of strong intersections between the spheres and the cells. Since there are r cells and for each cell there are only $O(k/r^{1/d})$ spheres strongly intersect it, it follows that

$$\sum_{v\in A}\sum_{i=1}^{t} s_i'(v) = O\Big(r\frac{k}{r^{1/d}}\Big) \le cntr^{(d-1)/d}, \tag{28}$$

for some constant c.

The sum $\sum_{v\in A}\sum_{i=1}^{t} s_i''(v)$ counts the total number of weak intersections between the spheres and the cells. To bound this number, notice that for a fixed point $v\in A$ and a fixed cell P, there are at most two spheres centered at v which intersect P weakly (if P weakly intersects S then either P is inside S or P is outside S). Thus we have

$$\sum_{v\in A}\sum_{i=1}^{t} s_i''(v) \le 2nr. \tag{29}$$

The estimates in (25–29) together imply that

$$\begin{aligned} M_r &\ge \frac{1}{m}\Big(\sum_{v\in A}\sum_{i=1}^{t} |A\cap S_i(v)| - cntr^{(d-1)/d} - 2nr\Big) \\ &= \frac{1}{m}\Big(n(n-1) - cntr^{(d-1)/d} - 2nr\Big). \end{aligned} \tag{30}$$

This, together with the upper bound (23), yields

$$\frac{n^2}{r} = \Omega\Big(\frac{1}{m}\Big(n(n-1) - cntr^{(d-1)/d} - 2nr\Big)\Big). \tag{31}$$

Let us choose $r = \epsilon(\frac{n}{t})^{d/(d-1)}$, where ϵ is a positive constant. Setting ϵ sufficiently small compared to $1/c$, we have that $cntr^{(d-1)/d} \le n^2/3$ and also

that $2nr \le n^2/6$ (the second inequality is due to the fact that $t = \Omega(n^{1/d})$, mentioned in the introduction). So with this setting of r, we have

$$n(n-1) - cntr^{(d-1)/d} - 2nr \ge n(n-1) - n^2/2 \ge n^2/3.$$

So with this choice of r, (31) implies

$$\frac{n^2}{\epsilon(\frac{n}{t})^{d/(d-1)}} = \Omega\left(\frac{n^2}{m}\right). \tag{32}$$

It follows that

$$t^{d/(d-1)} = \Omega\left(\frac{n^{d/(d-1)}}{m}\right), \tag{33}$$

namely,

$$t = \Omega\left(\frac{n}{m^{(d-1)/d}}\right), \tag{34}$$

concluding the proof. ∎

5. Proof of Theorem 2.2

This proof, in spirit, is very similar to the previous one. The main (and only) difference here is that we now consider triplets instead of pairs. We only need to show that

$$t(A) = \Omega\left(\frac{n^{(d+1)/2d}}{m^{(d-1)/2d}}\right).$$

We call a triplet in A *consistent* if its three elements belong to the same cell. Let N_r denote the number of consistent triplets. Similar to the previous proof, we are going to estimate N_r from both above and below.

Since $|P_i \cap A| = O(n/r)$ for all $1 \le i \le r$,

$$N_r = O\left(r\binom{n/r}{3}\right) = O\left(\frac{n^3}{r^2}\right). \tag{35}$$

To lower bound N_r, again let us consider a point $v \in A$. Consider a sphere $S_i(v)$. Without loss of generality, we can assume that the cells intersecting $S_i(v)$ are $P_1, \dots, P_l$. Let $x_j = |P_j \cap S_i(v)|$, $1 \le j \le l$. The number of consistent triplets on S_i is

$$\sum_{j=1}^{l}\binom{x_j}{3} \ge \sum_{j=1}^{l}(x_j - 2) = |A \cap S_i(v)| - 2s_i(v). \tag{36}$$

Summing the above estimate over all spheres $S_i(v)$'s and then summing over all $v \in A$ give us

$$\sum_{v\in A} \sum_{S_i(v)} (|A \cap S_i(v)| - 2s_i(v))$$

consistent triplets. Similar to the previous proof, this is not yet an estimate for N_r, as a triplet can be counted many times. Notice that if the three vertices of a consistent triplet T are colinear, then there is no point which is at the same distance from the vertices of T. Otherwise, the points which are at the same distance from the vertices of T lie on a hyperplane of co-dimension 2. By the assumption of the theorem, a hyperplane of co-dimension 2 contains at most m points from A, so the multiplicity of T is at most m. It follows that

$$N_r \geq \frac{1}{m} \sum_{v\in A} \sum_{S_i(v)} (|A \cap S_i(v)| - 2s_i(v)). \tag{37}$$

The estimates in (25–29) from the previous proof imply that

$$\begin{aligned} N_r &\geq \frac{1}{m}\left(\sum_{v\in A} \sum_{i=1}^{t} |A \cap S_i(v)| - cntr^{(d-1)/d} - cnr \right) \\ &= \frac{1}{m}\Big(n(n-1) - cntr^{(d-1)/d} - cnr\Big), \end{aligned} \tag{38}$$

for some constant c. This, together with the upper bound (35), yields

$$\frac{n^3}{r^2} = \Omega\Big(\frac{1}{m}\Big(n(n-1) - cntr^{(d-1)/d} - cnr\Big)\Big). \tag{39}$$

We set r as before: $r = \epsilon(\frac{n}{t})^{d/(d-1)}$, where ϵ is a small positive constant. With this choice of r, (39) implies

$$\frac{n^3}{\epsilon^2(\frac{n}{t})^{2d/(d-1)}} = \Omega\left(\frac{n^2}{m}\right). \tag{40}$$

It follows that

$$t^{2d/(d-1)} = \Omega\left(\frac{n^{(d+1)/(d-1)}}{m}\right), \tag{41}$$

namely,

$$t = \Omega\left(\frac{n^{(d+1)/2d}}{m^{(d-1)/2d}}\right), \tag{42}$$

concluding the proof. ∎

dc_52_10

6. Concluding remarks

Distinct distances in homogeneous sets. A set A of cardinality n is homogeneous if it is a subset of a full dimensional hypercube of volume n and any unit hypercube contains only $O(1)$ elements of A. If A is homogeneous, then a hyperplane of co-dimension 1 contains only $O(n^{(d-1)/d})$ elements of A. Thus, in Theorem 2.1, we can set $m = n^{(d-1)/d}$ to get

$$t(A) = \Omega\left(n^{2/d-1/d^2}\right), \tag{43}$$

for any $d \geq 3$. This estimate improves a results of Iosevich [8,9], who used a stronger definition of homogeneity. Applying Theorem 2.2 instead of Theorem 2.1 results in the same bound. For the special case $d = 3$, we can obtain the bound $\Omega(n^{.5794})$ (see [13]) for details. The homogeneity assumption is very popular among analysts, since their finite sets are usually the discretized versions of continuous domains.

Acknowledgement. The first author is supported by NSERC and OTKA grants. The second author is supported by an A. Sloan fellowship, an NSF Career Award and NSF grant DMS-0200357.

References

[1] J. Pach and P. Agarwal: *Combinatorial geometry*, Wiley-Interscience Series in Discrete Mathematics and Optimization, A Wiley-Interscience Publication, John Wiley & Sons, Inc., New York, 1995, xiv+354 pp.

[2] B. Aronov, J. Pach, M. Sharir and G. Tardos: Distinct Distances in Three and Higher Dimensions, *Combinatorics, Probability and Computing* **13(3)** (2004), 283–293.

[3] B. Chazelle and J. Friedman: A deterministic view of random sampling and its use in geometry, *Combinatorica* **10(3)** (1990), 229–249.

[4] F. Chung: The number of different distances determined by n points in the plane, *J. Combin. Theory Ser. A* **36(3)** (1984), 342–354.

[5] F. Chung, E. Szemerédi and W. Trotter: The number of different distances determined by a set of points in the Euclidean plane, *Discrete Comput. Geom.* **7(1)** (1992), 1–11.

[6] K. Clarkson, H. Edelsbrunner, L. Gubias, M. Sharir and E. Welzl: Combinatorial complexity bounds for arrangements of curves and spheres, *Discrete Comput. Geom.* **5** (1990), 99–160.

[7] P. Erdős: On sets of distances of n points, *Amer. Math. Monthly* **53** (1946), 248–250.

[8] A. Iosevich: Curvature, Combinatorics, and the Fourier Transform; *Notices of the American Mathematical Society* **48** (2001), 577–583.

[9] A. Iosevich: Szemerédi–Trotter incidence theorem, related results, and amusing consequences; in *Proceedings of Minicorsi di Analisi Matematica, Padova* (to appear).

dc_52_10

[10] J. Matousek: *Lectures on Discrete Geometry*, Graduate Texts in Mathematics, **212**, Springer-Verlag, New York, 2002, xvi+481 pp.
[11] L. Moser: On the different distances determined by n points, *Amer. Math. Monthly* **59** (1952), 85–91.
[12] J. Solymosi and Cs. D. Tóth: Distinct distances in the plane, *Discrete Comput. Geom.* **25(4)** (The Micha Sharir birthday issue) (2001), 629–634.
[13] J. Solymosi and V. H. Vu: Distinct distances in high dimensional homogeneous sets, in *Towards a Theory of Geometric Graphs (J. Pach, ed.)*, pp. 259–268, Contemporary Mathematics, vol. **342**, Amer. Math. Soc., 2004.
[14] L. Székely: Crossing numbers and hard Erdős problems in discrete geometry, *Combin. Probab. Comput.* **6(3)** (1997), 353–358.
[15] G. Tardos: On distinct sums and distinct distances, *Advances in Mathematics* **180(1)** (2003), 275–289.

József Solymosi

Department of Mathematics
UBC
Vancouver, BC, V6T 1Z2
Canada
solymosi@math.ubc.ca

Van H. Vu

Department of Mathematics
Rutgers
Piscataway, NJ 08854-8019
USA
vanvu@math.rutgers.edu

Discrete Comput Geom 35:537–549 (2006)
DOI: 10.1007/s00454-006-1232-4

Discrete & Computational
Geometry

Distinct Distances in Homogeneous Sets in Euclidean Space*

József Solymosi[1] and Csaba D. Tóth[2]

[1]Department of Mathematics, University of British Columbia,
Vancouver, British Columbia, Canada V6T 1Z2
solymosi@math.ubc.ca

[2]Department of Mathematics, Massachusetts Institute of Technology,
Cambridge, MA 02139, USA
toth@math.mit.edu

Abstract. It is shown that every homogeneous set of n points in d-dimensional Euclidean space determines at least $\Omega(n^{2d/(d^2+1)}/\log^{c(d)} n)$ distinct distances for a constant $c(d) > 0$. In three-space the above general bound is slightly improved and it is shown that every homogeneous set of n points determines at least $\Omega(n^{0.6091})$ distinct distances.

1. Introduction

The history of the *distinct distance problem* goes back to Erdős [10] who asked the question: What is the minimal number $g_d(n)$ of distinct distances determined by n points in d-dimensional Euclidean space $\mathbb{R}^d$? n points in the d-dimensional integer grid $[1, 2, \ldots, n^{1/d}]^d$ show that $g_d(n) = O(n^{2/d})$ for any $d \geq 2$ and, in particular, $g_2(n) = O(n/\sqrt{\log n})$. Erdős conjectured that these bounds are essentially optimal [11].

An initial lower bound of $g_2(n) \geq \Omega(\sqrt{n})$ by Erdős [10] was improved over the last almost 60 years by Moser, Beck, Chung, Szemerédi, Trotter, and Székely [19], [3], [5], [6], [25]. Research efforts have lead to several powerful methods (such as the crossing theory [25] and the ε-cutting theory [7]) which, in turn, found innumerable applications in discrete and computational geometry. An excellent survey by Pach and Sharir [20] elaborate on the history of the distinct distance problem and its connections to other fields of discrete mathematics. Determining the order of magnitude of $g_2(n)$ (and $g_d(n)$ for every $d \in \mathbb{N}$) seems elusive. The currently known best lower bound in the plane,

* The research by József Solymosi was supported by OTKA and NSERC grants.

$g_2(n) = \Omega(n^{0.8641})$, is due to Katz and Tardos [17]. Their proof combines a method of Solymosi and Tóth [21] with results from entropy and additive number theory.

Not much work has been done in higher dimensions. After some initial results by Clarkson et al. [7] and by Spencer et al. [24], Aronov et al. [2] have showed recently that the number of distinct distances determined by a set of n points in three-dimensional space is $g_3(n) = \Omega(n^{77/141-\varepsilon}) = \Omega(n^{0.5460})$ for any $\varepsilon > 0$. Solymosi and Vu [23] proved a general lower bound of $g_d(n) = \Omega(n^{2/d-2/d(d+2)})$ for any fixed $d \geq 4$.

In this paper we consider the minimum number $h_d(n)$ of distinct distances in homogeneous sets of n points in $\mathbb{R}^d$. A finite point set $P \subset \mathbb{R}^d$ is *homogeneous* if the following two conditions hold: P lies in the interior of an axis-aligned d-dimensional cube C of volume $|P|$, and any unit cube in $\mathbb{R}^d$ contains at most $O(1)$ points of P. Homogeneous sets represent an important special case for the distinct distance problem because the best known upper bound constructions (the d-dimensional integer grids) are homogeneous, and because of numerous connections to harmonic analysis [4], [12], [14], [16], [18]. Iosevich [13] studied the distinct distance problem for homogeneous sets (with additional restrictions). He showed that $h_d(n) = \Omega(n^{3/2d})$ for any fixed $d \geq 2$. Solymosi and Vu [22] proved a general bound of $h_d(n) = \Omega(n^{2/d-1/d^2})$ for every dimension $d \geq 2$. For $d = 3$, they have also obtained a slightly better bound $h_3(n) = \Omega(n^{0.5794})$. In this paper we improve all previous lower bounds on the number of distinct distances in homogeneous sets of n points in $\mathbb{R}^d$, $d \geq 3$.

Theorem 1. *For every $d \in \mathbb{N}$, there is a constant c_d such that in every homogeneous set P of n points in $\mathbb{R}^d$, there is a point $p \in P$ from which there are at least*

$$c_d\, n^{2d/(d^2+1)} \log^{(1-d^2)/(d^2+1)} n$$

distinct distances measured to other points of P. In particular, we have $h_d(n) \geq c_d\, n^{2d/(d^2+1)} \log^{(1-d^2)/(d^2+1)} n$.

For $d = 3, 4$, and 5, our general lower bound is $h_3(n) = \Omega(n^{0.5999})$, $h_4(n) = \Omega(n^{0.4705})$, and $h_5(n) = \Omega(n^{0.3846})$. In three dimensions we slightly improve on this bound and prove the following:

Theorem 2. *In every homogeneous set P of n points in $\mathbb{R}^3$, there is a point $p \in P$ from which there are at least*

$$\Omega(n^{53/87}) = \Omega(n^{0.6091})$$

distinct distances measured to other points of P. In particular, we have $h_3(n) = \Omega(n^{53/87})$.

We prove Theorem 1 in Section 3. The proof of Theorem 2 can be found in Section 4. In the next section we present a key lemma on the number of k-flats incident to many points in a homogeneous point set in $\mathbb{R}^d$, for $1 \leq k < d$.

2. Rich Hyperplanes in Homogeneous Sets

Consider a set P of n points in $\mathbb{R}^d$. We say that a k-flat (a k-dimensional affine subspace) is *m-rich* if it is incident to at least m points of P. The celebrated Szemerédi–Trotter Theorem [26] states that for n points in the plane, the number of m-rich lines (1-flats) is at most $O(n^2/m^3 + n/m)$, and this bound is tight in the worst case.

The number of m-rich k-flats in $\mathbb{R}^d$ has been intensely studied. The Szemerédi–Trotter type results have widespread applications in discrete and combinatorial geometry. The Szemerédi–Trotter Theorem's multi-dimensional generalizations [1], [8], [9] always impose some kind of restriction on the point set or on the set of k-flats, otherwise m points on a line give rise to infinitely many m-rich k-flats for any $2 \leq k \leq d$.

We adopt the following terminology. A set of $k+1$ points in $\mathbb{R}^d$, $k \leq d$, is *affine independent* if it is contained in a unique k-flat, which is said to be *span*ned by the point set. A point set P *determines* all the k-flats spanned by some $k+1$ affine independent points of P. For a constant $\alpha > 0$, a finite point set $P \subset \mathbb{R}^d$ that spans a k-flat is *α-degenerate* if any $(k-1)$-flat contains at most $\alpha \cdot |P|$ points of P. For a finite point set $P \subset \mathbb{R}^d$ and a constant $\alpha > 0$, we say that a k-flat F is *α-degenerate* if the point set $P \cap F$ is α-degenerate. Note, for example, that all points of $P \cap F$ in a 1-degenerate k-flat F may lie on a $(k-1)$-flat, but an α-degenerate k-flat for $\alpha < 1$ must be spanned by points of P. We recall a result of Beck [3] on α-degenerate hyperplanes.

Theorem 3 [3]. *For every $k \in \mathbb{N}$, there are constants $\alpha_k, \beta_k > 0$ with the following property. For every $d \in \mathbb{N}$ and every finite point set $P \subset \mathbb{R}^d$, if a k-flat F is α_k-degenerate, then $P \cap F$ spans at least $\beta_k \cdot |F \cap P|^k$ distinct $(k-1)$-flats.*

Elekes and Tóth [9] proved that for every dimension $d \in \mathbb{N}$, there is a constant $\gamma_d > 0$ such that the number of m-rich γ_d-degenerate hyperplanes for n points in $\mathbb{R}^d$ is at most $O(n^d/m^{d+1} + n^{d-1}/m^{d-1})$. The first term, $O(n^d/m^{d+1})$, is dominant only if $m = O(\sqrt{n})$. We show below a much stronger upper bound *for homogeneous sets*. A homogeneous set of n points in $\mathbb{R}^d$ determines at most $O(n^d/m^{d+1})$ distinct m-rich hyperplanes for every $m \in \mathbb{N}$, $d \leq m \leq n$.

We formulate our result for a slightly more general class of point sets, where n denotes the volume of the enclosing cube, rather than the number of points. We say that a point set P is *well separated* if any unit cube in $\mathbb{R}^d$ contains at most $O(1)$ points of P. By definition, every homogeneous set of n points in $\mathbb{R}^d$ is well separated, and lies in a cube of volume n.

Let $f_{d,k}(P, m)$ denote the maximal number of m-rich k-flats in a well separated point set P contained in the interior of a d-dimensional cube of volume n in $\mathbb{R}^d$, and let

$$f_{d,k}(n, m) = \max_{P \subset \mathbb{R}^d, |P|=n} f_{d,k}(P, m).$$

Solymosi and Vu [22] established the following lemma for the number of m-rich lines in homogeneous sets of n points in $\mathbb{R}^d$. Their proof carries over verbatim for well separated sets of volume n.

Lemma 4 [22]. *For every $d \in \mathbb{N}$, there is a constant c_d such that*

$$f_{d,1}(n,m) \leq c_d \frac{n^2}{m^{d+1}}.$$

We extend their result for arbitrary $k \in \mathbb{N}$, $1 \leq k \leq d-1$.

Lemma 5. *For every $d, k \in \mathbb{N}$, $1 \leq k < d$, there is a constant $c_{d,k}$ such that*

$$f_{d,k}(n,m) \leq c_{d,k} \frac{n^{k+1}}{m^{d+1}}.$$

The example of the d-dimensional integer grid $[1, 2, \ldots, n^{1/d}]^d$ shows that this bound is best possible for every $m \in \mathbb{N}$, $1 \leq m \leq n^{k/d}$.

Proof. For a fixed $d \in \mathbb{N}$, we prove that $f_{d,k}(n,m) = O(n^{k+1}/m^{d+1})$. We proceed by induction on k, $1 \leq k \leq d$. The base case, $k = 1$, is equivalent to Lemma 4. Assume that $1 < k \leq d$ and that $f_{d,k_0}(n_0,m) = O(n_0^{k_0+1}/m^{d+1})$ for every k_0, $1 \leq k_0 < k$, and $n_0 \in \mathbb{N}$.

Consider a well separated set P that lies in the interior of a d-dimensional cube C of volume n. Clearly, we have $|P| = O(n)$. We may choose an orthogonal coordinate system such that all coordinates of every point of P are irrational and P lies in the interior of cube C, whose vertices have rational coordinates. This guarantees that for any subdivision of C into congruent subcubes, every point of P lies in the *interior* of a subcube. For $i = 0, 1, \ldots, \lceil \log n^{1/d} \rceil$, let $\mathcal{C}_i$ denote the subdivision of the cube C into 2^{id} congruent cubes. For instance, $\mathcal{C}_0 = \{C\}$, $\mathcal{C}_1$ is a subdivision of C into 2^d cubes, and $\mathcal{C}_{\lceil (\log n)/d \rceil}$ is a subdivision into constant volume cubes. There is a constant $\delta_d > d$ such that every k-flat F intersects at most $\delta_d |\mathcal{C}_i|^{k/d} = \delta_d 2^{ik}$ cubes of $\mathcal{C}_i$. If we put

$$\mu = \left\lfloor \frac{1}{k} \log \frac{m}{4\delta_d(k+1)} \right\rfloor,$$

then every m-rich k-flat F is incident to an average of at least $m/(\delta_d 2^{\mu k}) \geq 4(k+1)$ points in a cube $Q \in C_\mu$. That is, at least $m/2$ points of $P \cap F$ lie in subcubes $Q \in C_\mu$ where $|P \cap F \cap Q| \geq 2(k+1)$.

Let α_k and β_k be the constants from Theorem 3. Let $\mathcal{F}$ denote the m-rich k-flats. We classify the k-flats in $\mathcal{F}$ as follows:

- $\mathcal{F}_1 = \{F \in \mathcal{F}\colon\ P \cap F$ is not α_k-degenerate$\}$,
- $\mathcal{F}_2 = \{F \in \mathcal{F}$: at least $m/4$ points of $P \cap F$ lie in cubes $Q \in C_\mu$ such that the point set $P \cap F \cap Q$ is α_k-degenerate$\}$,
- $\mathcal{F}_3 = \mathcal{F} \backslash (\mathcal{F}_1 \cup \mathcal{F}_2)$.

We show below that $|\mathcal{F}_q| = O(n^{k+1}/m^{d+1})$, for $q = 1, 2$, and 3. Every $F \in \mathcal{F}_1$ contains an $(\alpha_k m)$-rich $(k-1)$-flat. By induction, the number of $(\alpha_k m)$-rich $(k-1)$-flats is $O(n^k/(\alpha_k m)^{d+1}) = O(n^k/m^{d+1})$. Every $(\alpha_k m)$-rich $(k-1)$-flat R can be extended to an m-rich k-flat in $O(n)$ different ways: R together with a point of $P \backslash R$ spans a k-flat. This gives an upper bound $|\mathcal{F}_1| = O(n^{k+1}/m^{d+1})$.

For an upper bound on $|\mathcal{F}_2|$, we consider the subdivision $\mathcal{C}_\mu$. Let K denote the affine independent $(k+1)$-element subsets of P that determine some m-rich k-flat in $\mathcal{F}_2$ and lie in some cube $Q \in \mathcal{C}_\mu$. The volume of every cube $Q \in \mathcal{C}_\mu$ is $O(n/2^{\mu d}) = O(n/m^{d/k})$. Since P is well separated, we have $|P \cap Q| = O(n/m^{d/k})$. A trivial upper bound for the number of affine independent $(k+1)$-element sets in all cubes of $\mathcal{C}_\mu$ is

$$|K| \leq |\mathcal{C}_\mu| \cdot \left(O\left(\frac{n}{m^{d/k}}\right)\right)^{k+1} = O\left(\frac{n^{k+1}}{m^d}\right).$$

We obtain a lower bound for $|K|$ by counting the affine independent sets in each $F \in \mathcal{F}_2$. At least $m/4$ points of $P \cap F$ lie in cubes $Q \in \mathcal{C}_\mu$ where the point set $P \cap F \cap Q$ is α_k-degenerate. By Theorem 3, every α_k-degenerate set $P \cap F \cap Q$ determines at least $\beta_k |P \cap F \cap Q|^{k+1}$ affine independent $(k+1)$-element sets. If we denote by $K(F)$ the number of $(k+1)$-element subsets of K that span F, then we have

$$|K(F)| \geq \sum_{\substack{Q \in \mathcal{C}_\mu \\ Q \cap F \neq \emptyset}} \beta_k |P \cap F \cap Q|^{k+1} \geq \delta_d 2^{\mu k} \left(\frac{m/4}{\delta_d 2^{\mu k}}\right)^{k+1} = \Omega(m^{k+1} 2^{-\mu k^2}) = \Omega(m).$$

We conclude that $|K| = \sum_{F \in \mathcal{F}_2} \Omega(m) = |\mathcal{F}_2| \cdot \Omega(m)$. By contrasting the upper and lower bounds for $|K|$, we get $|\mathcal{F}_2| = O(n^{k+1}/m^{d+1})$.

Finally, we consider $\mathcal{F}_3$. For every m-rich k-flat $F \in \mathcal{F}_3$, we define a set $S(F)$ of cubes from $\mathcal{C}_i$, $i = 1, 2, \ldots, \log n^{1/d}$. A cube $Q \in \mathcal{C}_i$ is in $S(F)$ if and only if the point set $P \cap F \cap Q$ is *not* α_k-degenerate, but $P \cap F \cap Q(i')$ is α_k-degenerate for every i', $0 \leq i' < i$, where $Q(i')$ is the (unique) cube $Q(i') \in C_{i'}$ containing Q. If $P \cap F$ is not α_k-degenerate, for example, then $C \notin S(F)$. Observe that the cubes of $S(F)$ are pairwise interior disjoint and they jointly cover $P \cap F \cap C$. We denote by $\dim(X)$ the dimension of the affine subspace spanned by a finite point set X. For each $F \in F_3$, we further classify the cubes in $S(F)$ according to three parameters: For $i \in \{1, 2, \ldots, \mu\}$, $j \in \{0, 1, \ldots, \log m\}$, and $r \in \{1, \ldots, k-1\}$, let $S(F, i, j, r)$ denote the set of cubes $Q \in S(F)$ such that

1. $Q \in \mathcal{C}_i$,
2. $2^{j-1} \cdot m/\delta_d 2^{ik} \leq |P \cap F \cap Q| < 2^j \cdot m/\delta_d 2^{ik}$,
3. $r = \min(k-1, \dim(P \cap F \cap Q))$.

Some of the cubes $Q \in S(F)$ are not included in any $S(F, i, j, r,) \subset S(F)$: This is the case for every $Q \in S(F) \cap \mathcal{C}_i$ for which $|P \cap F \cap Q| < (m/\delta_d 2^{ik+1})$ or $\mu < i$. The total of number points of $P \cap F$ in these cubes is less than

$$\sum_{\substack{Q \in S(F) \cap \mathcal{C}_i \\ 0 < i < \mu}} |P \cap F \cap Q| + \sum_{\substack{Q \in S(F) \cap \mathcal{C}_i \\ i \geq \mu}} \frac{m}{\delta_d 2^{ik}} < \frac{m}{2} + \frac{m}{4} = \frac{3m}{4}.$$

Therefore, the cubes in $S(F, i, j, r)$ for all i, j, r jointly contain at least $m/4$ points of $P \cap F$:

$$\sum_{i=1}^{\mu} \sum_{j=0}^{\log m} \sum_{r=1}^{k-1} |S(F, i, j, r)| \cdot \frac{2^{j-1} m}{\delta_d 2^{ik}} \geq \frac{m}{4}. \tag{1}$$

For every $Q \in S(F, i, j, r)$, there is an r-flat $R \subset F$, such that $|P \cap R \cap Q| \geq \alpha_k |P \cap F \cap Q| \geq \alpha_k 2^{j-1} m/(\delta_d 2^{ik}) = \Theta(2^{j-ik} m)$. Let us denote by Q' the cube in $\mathcal{C}_{i-1}$ that contains $Q \in \mathcal{C}_i$. Since $P \cap F \cap Q'$ is already α_k-degenerate, we have $|P \cap R \cap Q| \leq \alpha_k |P \cap F \cap Q'|$. Let $D(Q, R)$ be the set of all $(k-r)$-element affine independent sets $u \subset (P \cap F \cap Q') \setminus R$ such that R and u together span F. Since $P \cap F \cap Q'$ is α_k-degenerate, there are $\Theta(|P \cap F \cap Q'|^{k-r})$ sets in $D(Q, R)$. Let $D'(Q, R)$ be a subset of $D(Q, R)$ of size $\Theta(|P \cap F \cap Q|^{k-r}) = \Theta((m2^{j-ik})^{k-r})$.

Let $T(F, i, j, r)$ denote the set of triples (Q, R, u) such that $Q \in S(F, i, j, r)$, R is an r-flat with $|P \cap R \cap Q| \geq \alpha_k |P \cap F \cap Q|$, and $u \in D'(Q, R)$. We have a lower bound

$$|T(F, i, j, r)| \geq |S(F, i, j, r)| \cdot \Theta((m2^{j-ik})^{k-r}).$$

Let us put

$$\tau(F, i, j, r) = \frac{|T(F, i, j, r)|}{(m2^{j-ik})^{k-r-1}},$$

and then inequality (1) can be rewritten as

$$\sum_{i=1}^{\mu} \sum_{j=0}^{\log m} \sum_{r=1}^{k-1} \tau(F, i, j, r) \geq \sum_{i=1}^{\mu} \sum_{j=0}^{\log m} \sum_{r=1}^{k-1} |S(F, i, j, r)| \cdot \Omega(m2^{j-ik}) \geq \Omega(m).$$

By summing over all $F \in \mathcal{F}_3$, we get

$$\sum_{F \in \mathcal{F}_3} \sum_{i=1}^{\mu} \sum_{j=0}^{\log m} \sum_{r=1}^{k-1} \tau(F, i, j, r) \geq |\mathcal{F}_3| \cdot \Omega(m). \tag{2}$$

We also compute an upper bound for the quantity on the left side of inequality (2). First, we give an upper bound on the number of triples $(Q, R, u) \in T(F, i, j, r)$ for all $F \in \mathcal{F}_3$. Recall that $(Q, R, u) \in T(F, i, j, r)$ implies that $Q \in \mathcal{C}_i$, and R is an r-flat incident to $\ell = \Omega(m2^{j-ik})$ points of $P \cap Q$. Every cube $Q \in \mathcal{C}_i$ has volume $n/2^{id}$ and $P \cap Q$ is well separated. By our induction hypothesis, the number of ℓ-rich r-flats in $P \cap Q$ is $O((n/2^{id})^{r+1}/\ell^{d+1})$. The cube $Q' \in C_{i-1}$ contains $|P \cap Q'| = O(n/2^{(i-1)k}) = O(n/2^{id})$ points. So $P \cap Q'$ contains $(O(n/2^{id}))^{k-r}$ distinct $(k-r)$-element subsets. For all $Q \in \mathcal{C}_i$, we obtain an upper bound

$$\sum_{F \in \mathcal{F}_3} |T(F, i, j, r)| \leq |\mathcal{C}_i| \cdot O\left(\frac{(n/2^{id})^{r+1}}{(m2^{j-ik})^{d+1}}\right) \cdot O\left(\left(\frac{n}{2^{id}}\right)^{k-r}\right),$$

$$\sum_{F \in \mathcal{F}_3} |T(F, i, j, r)| \leq O\left(\frac{n^{k+1}}{m^{d+1}} \cdot 2^{ik-j(d+1)}\right). \tag{3}$$

After dividing by $(m2^{j-ik})^{k-r-1}$, we sum inequality (3) over all i, j, and r:

$$\sum_{F \in \mathcal{F}_3} \tau(F, i, j, r) \leq O\left(\frac{n^{k+1}}{m^{d+1}} \cdot 2^{ik-j(d+1)} \cdot \left(\frac{2^{ik}}{2^j m}\right)^{k-r-1}\right),$$

$$\sum_{r=1}^{k-1} \sum_{F \in \mathcal{F}_3} \tau(F, i, j, r) \leq O\left(\frac{n^{k+1}}{m^{d+1}} \cdot 2^{ik-j(d+1)}\right),$$

$$\sum_{j=0}^{\log m}\sum_{r=1}^{k-1}\sum_{F\in\mathcal{F}_3}\tau(F,i,j,r) \;\le\; O\left(\frac{n^{k+1}}{m^{d+1}}\cdot 2^{ik}\right),$$

$$\sum_{i=1}^{\mu}\sum_{j=0}^{\log m}\sum_{r=1}^{k-1}\sum_{F\in\mathcal{F}_3}\tau(F,i,j,r) \;\le\; O\left(\frac{n^{k+1}}{m^{d+1}}\cdot m\right). \tag{4}$$

By contrasting inequalities (2) and (4), we conclude that $|\mathcal{F}_3| = O(n^{k+1}/m^{d+1})$. This completes the proof of Lemma 5. □

Corollary 6. *For every $d, k \in \mathbb{N}$, $1 \le k < d$, the number of incidences of points and m-rich k-flats in a homogeneous set of n points in $\mathbb{R}^d$ is at most*

$$O\left(\frac{n^{k+1}}{m^d}\right).$$

Proof. In any homogeneous point set of size n in $\mathbb{R}^d$, the number of incidences of points and m-rich k-flats is bounded by

$$mf_{d,k}(P,m) + \sum_{j=m+1}^{n} f_{d,k}(P,j) \le O\left(\frac{n^{k+1}}{m^d}\right) + \sum_{j=m+1}^{n} O\left(\frac{n^{k+1}}{j^{d+1}}\right) \le O\left(\frac{n^{k+1}}{m^d}\right). \;\square$$

3. Proof of Theorem 1

We are given a homogeneous set P of n points in d-dimensions. We may choose an orthogonal coordinate system such that all coordinates of every point of P are irrational and P lies in the interior of cube C, whose vertices have rational coordinates. This guarantees that for any subdivision of C into congruent subcubes, every point of P lies in the *interior* of a subcube. Let t denote the maximum number of distinct distances measured from a point of P (including distance 0). There is a constant $\delta_d > d$ such that for every $s \in \mathbb{N}$, every hyperplane or sphere intersects the interior of at most $\delta_d s^{d-1}$ cubes in the subdivision of C into s^d congruent cubes. We subdivide C into s^d congruent subcubes $C_1, C_2, \ldots, C_{s^d}$, where

$$s = \left\lfloor \left(\frac{n}{4\delta_d t}\right)^{1/(d-1)} \right\rfloor.$$

Let T be a set of triples $(p,q,c) \in P^3$ such that

(i) $p \ne q$,
(ii) p and q lie in the same subcube C_i for some $1 \le i \le s^d$,
(iii) p and q are equidistant from c.

All points are located on nt spheres centered at the n points of P. The cubes C_i, $1 \le i \le s^d$, subdivide each sphere into *patches*. Since every sphere intersects at most $\delta_d s^{d-1}$ subcubes C_i, there are at most $\delta_d nts^{d-1} = n^2/4$ patches, where each patch lies entirely in a subcube C_i. There are n^2 sphere-point incidences. The average number of

points on a patch is at least four. If x points lie on a sphere patch centered at c, then this patch contributes $\binom{x}{2}2!$ triples (p, q, c) to T. We conclude that the number of triples is $|T| \geq \Omega(n^2)$.

For every $m \in \mathbb{N}$, let T_m denote the set of triples $(p, q, c) \in T$ such that the bisector hyperplane of the segment pq is incident to at least m but less than $2m$ points of P. Since every bisector plane is incident to less than n points, we can partition T into $\log n$ subsets

$$T = \bigcup_{j=0}^{\log n} T_{2^j}.$$

There is a value $m = 2^j$ for some $0 \leq j \leq \log n$, such that $|T_m| \geq |T|/\log n \geq \Omega(n^2/\log n)$.

For a pair $(p, q) \in P^2$, $p \neq q$, all points of the set $M(p, q) = \{c \in P\colon \operatorname{dist}(p, c) = \operatorname{dist}(q, c)\}$ lie on the bisector hyperplane of the line segment pq. Every bisector hyperplane intersects at most $\delta_d s^{d-1}$ subcubes, and in each subcube C_i it can bisect at most $|C_i \cap P|/2$ point pairs. So the number of pairs $(p, q) \in P^2$ bisected by the same hyperplane is at most

$$\delta_d s^{d-1} \cdot O\left(\frac{n}{s^d}\right) = O\left(\frac{n}{s}\right).$$

Let B_m denote the set of all bisector hyperplanes that bisect the pair (p, q) for some $(p, q, c) \in T_m$. By definition, every hyperplane in B_m is incident to at least m but less than $2m$ points of P. By Lemma 5, we have

$$|B_m| \leq O\left(\frac{n^d}{m^{d+1}}\right).$$

We can now give an upper bound for $|T_m|$. In a triple $(p, q, c) \in T_m$, point c lies on a bisector hyperplane of B_m. Each bisector hyperplane is incident to less than $2m$ points of P and bisects at most $O(n/s)$ pairs (p, q). Therefore

$$\Omega\left(\frac{n^2}{\log n}\right) \leq |T_m| \leq O\left(\frac{n^d}{m^{d+1}}\right) \cdot 2m \cdot O\left(\frac{n}{s}\right),$$

$$m^d \leq O\left(\frac{n^{d-1}\log n}{s}\right),$$

$$m \leq O\left(\frac{n^{(d-1)/d}\log^{1/d} n}{s^{1/d}}\right). \tag{5}$$

We obtain another upper bound for $|T_m|$ by the following argument. In a triple $(p, q, c) \in T_m$, both p and q lie in the same subcube $C_i \subset C$. There are s^d subcubes, and each subcube contains $(O(n/s^d))^2 \leq O(n^2/s^{2d})$ point pairs. Hence, there are at most $s^d \cdot O(n^2/s^{2d}) = O(n^2/s^d)$ such pairs (p, q). For each pair (p, q), where $(p, q, c) \in T_m$, there are at most $2m$ points $c \in P$ on the bisector hyperplane of pq. We conclude that

$$\Omega\left(\frac{n^2}{\log n}\right) \leq |T_m| \leq O\left(\frac{n^2}{s^d}\right) \cdot 2m.$$

Using the upper bound for m from inequality (5), we have

$$\begin{aligned} s^{(d^2+1)/d} &\leq O(n^{(d-1)/d} \cdot \log^{(d+1)/d} n), \\ \left(\frac{n}{t}\right)^{(d^2+1)/d(d-1)} &\leq O(n^{(d-1)/d} \cdot \log^{(d+1)/d} n), \\ \Omega(n^{2/(d-1)} \log^{-(1+d)/d} n) &\leq t^{(d^2+1)/d(d-1)}, \\ \Omega(n^{2d/(d^2+1)} \log^{(1-d^2)/(d^2+1)} n) &\leq t, \end{aligned}$$

as required. This completes the proof of Theorem 1. □

4. Proof of Theorem 2

Consider a homogeneous set P of n points in $\mathbb{R}^3$. Similarly to the previous section, we assume that all coordinates of every point in P are irrational, and the vertices of the bounding cube C have rational coordinates. Let t denote the maximum number of distinct distances measured from a point of P (including distance 0). We subdivide C into s^3 congruent cubes $C_1, C_2, \ldots, C_{s^3}$, for

$$s = \left\lfloor \sqrt{\frac{n}{\gamma t}} \right\rfloor,$$

where $\gamma > 0$ is a constant to be specified later.

By Theorem 3, $P \subset \mathbb{R}^3$ contains $\Omega(n^2)$ affine independent point pairs. This implies that there is a subset $P_0 \subset P$ such that $|P_0| \geq \Omega(n)$ and every $c \in P_0$ is incident to $\Omega(n)$ distinct lines spanned by P. For every $c \in P_0$, let $P(c) \subset P \backslash \{c\}$ be a set of $\Omega(n)$ points such that the lines cp, $p \in P(c)$, are distinct. For every point $c \in P_0$, let H_c be a unit sphere centered at c. For every $x \in \mathbb{R}^3 \backslash \{c\}$, we denote by $\hat{x}$ the projection of x to the unit sphere H_c. Points of $P(c)$ have distinct images in H_c under this projection. The set of projection points is denoted by

$$\hat{P}(c) := \{\hat{p} \colon c \in P(p)\}.$$

We partition the unit sphere H_c into $6s^2$ convex spherical regions $S_1(c), S_2(c), \ldots, S_{6s^2}(c)$ by $6s-12$ circular arcs: Consider an axis-parallel cube centered at c and subdivide each of its six faces into s^2 congruent squares, then project these squares to the sphere H_c from c. The area of each spherical region is $\Theta(1/s^2)$ and each one is contained in a disk of area $\Theta(1/s^2)$. Every circle on the sphere H_c intersects at most $O(s)$ regions. We then subdivide $\mathbb{R}^d \backslash \{c\}$ into $6s^2$ regions $R_i(c)$, $i = 1, 2, \ldots, 6s^2$, such that

$$R_i(c) = \{x \in \mathbb{R}^d \backslash \{c\} \colon \hat{x} \in S_i(c)\}.$$

For every $c \in P_0$ and $j = 1, 2, \ldots, 6s^2$, the region $R_j(c)$ contains $|P \cap R_i(c)| = O(n/s^2)$ points because the region $R_j(c) \cap C$ can be covered by $O(n/s^2)$ unit cubes. Note also that every plane incident to c intersects at most $O(s)$ regions $R_j(c)$, since every great circle of S intersects at most $O(s)$ spherical regions S_j. If F is a plane, then

$|F \cap R_j(c) \cap P| = O(n^{2/3}/s)$ because $F \cap C$ can be covered by $O(n^{2/3})$ unit cubes, and $\text{area}(F \cap R_j(c)) \leq O(\text{area}(F \cap C)/s)$.

For every $c \in P_0$, consider the at most t spheres centered at c that contain all points of $P(c)$. Every sphere S centered at c is partitioned into *patch*es by the cubes C_i, $1 \leq i \leq s^3$, and the regions $R_j(c)$, $1 \leq j \leq 6s^2$. We can partition C into the subcubes C_i, $1 \leq i \leq s^3$, by $3(s-1)$ planes. These planes partition every sphere S along $3(s-1)$ circles. Hence every sphere S is partitioned by $O(s)$ circular arcs into $O(s^2)$ patches. We partition the points of P lying on a patch into disjoint triples, after deleting at most two points from each patch if necessary. This produces a set Q of quadruples $(p, q, r, c) \in P^3 \times P_0$ such that

(i) the points p, q, and r are in $P(c)$;
(ii) p, q, and r lie on a sphere centered at c;
(iii) p, q, and r lie in the same subcube C_i for some $1 \leq i \leq s^3$;
(iv) p, q, and r lie in the same regions $R_j(c)$, for some $1 \leq j \leq 6s^2$;
(v) if $(p_1, q_1, r_1, c) \in Q$ and $(p_2, q_2, r_2, c) \in Q$, then $\{p_1, q_1, r_1\} \cap \{p_2, q_2, r_2\} = \emptyset$.

We give a lower bound on the number of quadruples in Q. Let $g(c)$ denote the number of patches on all $O(t)$ spheres centered at c: We have $g(c) = O(ts^2) = O(n/\gamma)$. The average number of points on a patch centered at c is $\Omega(\gamma n/g(c)) = \Omega(\gamma)$. We choose the constant $\gamma > 0$ such that a patch contains at least six points of $P(c)$ on average. If the kth patch contains a set of points $G_k(c) \subset P(c)$, then Q contains $\lfloor |G_k(c)|/3 \rfloor$ quadruples (p, q, r, c). We conclude that the total number of quadruples is

$$|Q| = \sum_{c \in P_0} \sum_{k=1}^{g(c)} \left\lfloor \frac{|G_k|}{3} \right\rfloor \geq \Omega\left(n \sum_{k=1}^{g(c)} (|G_k| - 2) \right) \geq \Omega(n^2).$$

We define the *multiplicity* of a pair $(p, q) \in P^2$ as

$$m(p, q) = |\{c \in P_0 \colon \exists r \text{ such that } (p, q, r, c) \in Q \text{ or } (q, r, p, c) \in Q \text{ or } (r, p, q, c) \in Q\}|.$$

We choose a parameter m to be specified later, and distinguish two types of quadruples in Q: A quadruple (p, q, r, c) is *low* if at least one edge of the triangle pqr has multiplicity at most m. A quadruple (p, q, r, c) is *high* if the multiplicity of all three edges of pqr are above m. Let Q^- and Q^+ denote the sets of low and high quadruples, respectively. We distinguish two cases: First we consider the case that $|Q^+| \leq |Q^-|$, then we proceed with the case $|Q^+| > |Q^-|$.

Case $|Q^+| \leq |Q^-|$. There are at least $\Omega(n^2)$ low quadruples in Q. We define a set of triples

$$T := \{(p, q, c) \colon (p, q, r, c) \in Q^-, m(p, q) \leq m\}.$$

We have extracted $|T| = \Omega(n^2)$ triples from Q^-. Similarly to the previous section, we compute an upper bound on $|T|$. Every pair (p, q) from a triple of T lies in one of the s^3 subcubes of C, and for every pair (p, q) there are at most m centers c. Therefore, we have an upper bound

$$|T| \leq s^3 \left(O\left(\frac{n}{s^3}\right) \right)^2 m = O\left(\frac{mn^2}{s^3}\right).$$

Comparing this upper bound with the lower bound $|T| = \Omega(n^2)$, we obtain

$$\begin{aligned} \Omega(s^3) &\leq m, \\ \Omega\left(\frac{n^{3/2}}{t^{3/2}}\right) &\leq m, \\ \Omega\left(\frac{n}{m^{2/3}}\right) &\leq t. \end{aligned} \tag{6}$$

Case $|Q^+| > |Q^-|$. At least half of the quadruples in Q are high, and so $|Q^+| \geq \Omega(n^2)$.

For every $c \in P_0$, project the points of $P(c)$ to the sphere H_c. If $(p, q, r, c) \in Q$, then the intersection of the bisector plane of pq and H_c is the *bisector* (great circle) of the segment $\hat{p}\hat{q}$ in the sphere H_c. A (possibly degenerate) triangle $\hat{p}\hat{q}\hat{r}$ defines three distinct bisectors. The bisectors of a triangle $\hat{p}\hat{q}\hat{r}$ meet in two antipodal points on the sphere. The triangles that determine the same triple of bisectors are similar (the center of similarity is the intersection of the bisectors). Specifically, if the triangles $\hat{p}_1\hat{q}_1\hat{r}_1$, $\hat{p}_2\hat{q}_2\hat{r}_2$, $\ldots$, $\hat{p}_\ell\hat{q}_\ell\hat{r}_\ell$ determine the same triple of bisectors, then the points $\hat{p}_1, \hat{p}_1, \ldots, \hat{p}_\ell$ are collinear (the points $\hat{q}_1, \hat{q}_2, \ldots, \hat{q}_\ell$ and $\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_\ell$ are also collinear). Every triple of bisectors determines a *family* of triangles. We define a *family of quadruples* to be a collection of quadruples $(p, q, r, c) \in Q^+$ with a common center c such that the triangles $\hat{p}\hat{q}\hat{r}$ form a family.

For every $c \in P_0$, we define a set of triangles in the sphere H_c by

$$T(c) = \{\hat{p}\hat{q}\hat{r}\colon\ (p, q, r, c) \in Q^+\}.$$

By construction, all these triangles have pairwise disjoint vertex sets. There is a set $P_1 \subseteq P_0$ of size $\Omega(n)$ such that for every $c \in P_1$, we have $|T(c)| = \Omega(n)$ triangles. For a point $c \in P_1$, let B_c denote the set of m-rich planes incident to c. We denote by $\hat{B}_c$ the set of intersections of planes in B_c and the unit sphere H_c, which are great circles on H_c. Note that the bisector of every edge $\hat{p}\hat{q}$ of a triangle of $T(c)$ is in $\hat{B}_c$.

For $c \in P_1$, we consider the partition of the sphere H_c into $6s^2$ regions $S_j(c)$, $1 \leq j \leq 6s^2$, defined above. Each triangle of $T(c)$ lies entirely in one of the regions. Let us denote by $T_j(c)$ the set of triangles of $T(c)$ in $S_j(c)$ for every $j = 1, 2, \ldots, 6s^2$. Since the triangles have disjoint vertex sets, we have $|T_j(c)| \leq |P \cap R_j(c)|/3 \leq O(n/s^2) = O(t)$. But $\sum_{j=1}^{6s^2} |T_j(c)| = \Omega(n)$, and so there are $\Omega(s^2)$ indices j such that $|T_j(c)| = \Omega(n/s^2) = \Omega(t)$. Vertices of similar triangles lie on three main circles. We have shown that every region $R_j(c)$ contains at most $O(n^{2/3}/s) = O(n^{1/6}t^{1/2})$ coplanar points. Hence, there are at least $\Omega(t^{1/2}/n^{1/6})$ families of triangles in $T_j(c)$. Since each such family determines three distinct bisectors of $\hat{B}(c)$, the triangles in $T_j(c)$ determine

$$\Omega\left(\left(\frac{t^{1/2}}{n^{1/6}}\right)^{1/3}\right) = \Omega\left(\frac{t^{1/6}}{n^{1/18}}\right)$$

distinct bisectors in $\hat{B}_c$. A bisector crosses at most $O(s)$ regions, and so we obtain the same bisector of $\hat{B}_c$ from at most $O(s)$ regions. We conclude that the number of bisectors determined by the $\Omega(n)$ triangles of $T(c)$ is

$$|B_c| \geq \frac{\Omega(s^2)}{O(s)} \cdot \Omega\left(\frac{t^{1/6}}{n^{1/18}}\right) \geq \Omega\left(\sqrt{\frac{n}{t}} \cdot \frac{t^{1/6}}{n^{1/18}}\right) = \Omega\left(\frac{n^{4/9}}{t^{1/3}}\right).$$

Each of the $\Omega(n)$ points of P_1 is incident to $\Omega(n^{4/9}/t^{1/3})$ distinct m-rich planes. This gives $\Omega(n^{13/9}/t^{1/3})$ incidences on m-rich planes of P. By Corollary 6, we have

$$\begin{aligned} \Omega\left(\frac{n^{13/9}}{t^{1/3}}\right) &\leq O\left(\frac{n^3}{m^3}\right), \\ m &\leq O(n^{14/27}t^{1/9}), \\ \Omega\left(\frac{m^9}{n^{14/3}}\right) &\leq t. \end{aligned} \tag{7}$$

In both cases we have derived lower bounds for t in terms of n and m. We choose $m \in \mathbb{N}$ such that we obtain the same result in both cases. By comparing inequalities (6) and (7), we have

$$\begin{aligned} \Omega\left(\frac{n^{3/2}}{t^{3/2}}\right) \leq m \leq O(n^{14/27}t^{1/9}), \\ \Omega(n^{53/87}) \leq t. \end{aligned} \tag{8}$$

The choice $m = n^{17/29}$ establishes inequality (8) in both cases. This completes the proof of Theorem 2. □

References

1. P. K. Agarwal and B. Aronov, Counting facets and incidences, *Discrete Comput. Geom.* **7** (1992), 359–369.
2. B. Aronov, J. Pach, M. Sharir, and G. Tardos, Distinct distances in three and higher dimensions, *Combin. Probab. Comput.* **13** (2004), 283–293.
3. J. Beck, On the lattice property of the plane and some problems of Dirac, Motzkin and Erdős, *Combinatorica* **3**(3–4) (1983), 281–297.
4. J. Bourgain, On the dimension of kayela sets and related maximal inequalities, *Geom. Funct. Anal.* **9** (1999), 256–282.
5. F. R. K. Chung, The number of different distances determined by n points in the plane, *J. Combin. Theory Ser. A* **36** (1984), 342–354.
6. F. R. K. Chung, E. Szemerédi, and W. T. Trotter, The number of different distances determined by a set of points in the Euclidean plane, *Discrete Comput. Geom.* **7** (1992), 1–11.
7. K. L. Clarkson, H. Edelsbrunner, L. J. Guibas, M. Sharir, and E. Welzl, Combinatorial complexity bounds for arrangements of curves and spheres, *Discrete Comput. Geom.* **5** (1990), 99–160.
8. H. Edelsbrunner and M. Sharir, A hyperplane incidence problem with applications to counting distances, in *Proc. SIGAL International Symposium on Algorithms* (T. Asano et al., eds.), vol. 450 of LNCS, Springer-Verlag, Berlin, 1990, pp. 419–428.
9. Gy. Elekes and Cs. D. Tóth, Incidences of not-too-degenerate hyperplanes, in *Proc. 21st ACM Symposium on Computational Geometry*, ACM Press, New York, 2005, pp. 16–21.
10. P. Erdős, On sets of distances of *n* points, *Amer. Math. Monthly* **53** (1946), 248–250.
11. P. Erdős, On some of my favourite theorems, in *Combinatorics*, *Paul Erdős is Eighty*, vol. 2 of Bolyai Society Mathematical Studies, János Bolyai Mathematical Society, Budapest, 1996, pp. 97–132.
12. S. Hofmann and A. Iosevich, Circular averages and Falconer–Erdős distance conjecture in the plane for random metrics, *Proc. Amer. Math. Soc.* **133**(1) (2005), 133–143.
13. A. Iosevich, Curvature, combinatorics, and the Fourier transform, *Notices Amer. Math. Soc.* **48** (2001), 577–583.
14. A. Iosevich, N. Katz, and S. Pedersen, Fourier basis and the Erdős distance problem, *Math. Res. Lett.* **6**(2) (1999), 251–255.

15. A. Iosevich and I. Łaba, Distance sets of well-distributed planar point sets, *Discrete Comput. Geom.* **31** (2004), 243–250.
16. N. H. Katz and T. Tao, Some connections between Falconer's distance set conjecture and sets of Furstenburg type, *New York J. Math.* **7** (2001), 149–187.
17. N. H. Katz and G. Tardos, A new entropy inequality for the Erdős distance problem, in *Towards a Theory of Geometric Graphs* (J. Pach, ed.), vol. 342 of Contemporary Mathematics, American Mathematical Society, Providence, RI, 2004, pp. 119–126.
18. S. Konyagin and I. Łaba, Distance sets of well-distributed planar sets for polygonal norms, *Israel J. Math.*, to appear.
19. L. Moser, On the different distances determined by n points, *Amer. Math. Monthly* **59** (1952), 85–91.
20. J. Pach and M. Sharir, Geometric incidences, in *Towards a Theory of Geometric Graphs* (J. Pach, ed.), vol. 342 of Contemporary Mathematics, American Mathematical Society, Providence, RI, 2004, pp. 185–223.
21. J. Solymosi and Cs. D. Tóth, Distinct distances in the plane, *Discrete Comput. Geom.* **25** (2001), 629–634.
22. J. Solymosi and V. Vu, Distinct distances in high dimensional homogeneous sets, in *Towards a Theory of Geometric Graphs* (J. Pach, ed.), vol. 342 of Contemporary Mathematics, American Mathematical Society, Providence, RI, 2004, pp. 259–268.
23. J. Solymosi and V. Vu, Near optimal bounds for the Erdős distinct distance problem in high dimensions, *Combinatorica*, to appear.
24. J. Spencer, E. Szemerédi, and W. T. Trotter, Unit distances in the Euclidean plane, in *Graph Theory and Combinatorics* (B. Bollobs, ed.), Academic Press, New York, 1984, pp. 293–303.
25. L. A. Székely, Crossing numbers and hard Erdős problems in discrete geometry, *Combin. Probab. Comput.* **6**(3) (1997), 353–358.
26. E. Szemerédi and W. T. Trotter Jr., Extremal problems in discrete geometry, *Combinatorica* **3**(3–4) (1983), 381–392.

Received March 21, 2005, *and in revised form April* 6, 2005, *and October* 17, 2005.
Online publication April 18, 2006.